

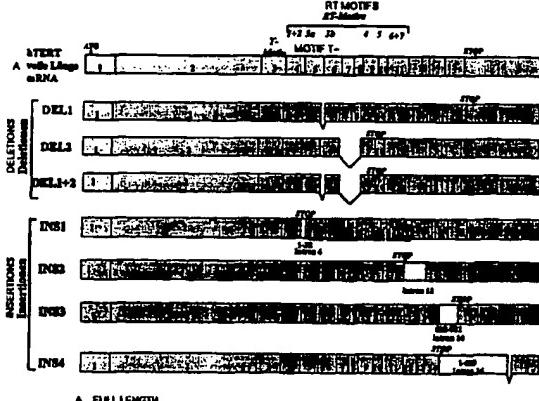
PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM

Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICH NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



(51) Internationale Patentklassifikation ⁶ : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027		A2	(11) Internationale Veröffentlichungsnummer: WO 99/33998 (43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)
(21) Internationales Aktenzeichen: PCT/EP98/08216 (22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98) (30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE (71) Anmelder (<i>für alle Bestimmungsstaaten ausser US</i>): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE). (72) Erfinder; und (75) Erfinder/Anmelder (<i>nur für US</i>): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Rogendorfstrasse 59, D-51061 Köln (DE). (74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D-51368 Leverkusen (DE).		(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIGO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Veröffentlicht <i>Ohne internationalem Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i>	
<p>(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF</p> <p>(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG</p> 			
<p>(57) Abstract</p> <p>The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.</p>			

(57) Zusammenfassung

Diese Erfundung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfundung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Oesterreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasiliens	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NB	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen Telomerase-Untereinheit und deren diagnostische und therapeutische Verwendung

5 Aufbau und Funktion der Chromosomenenden

- Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten 10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).
- 15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.
- 20 Des Weiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).
- 25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der 30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

15

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

20

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen
5 auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens
15 (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et*
20 *al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt
25 werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

- 5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein.
10 Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.
15

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweligen Strahlen schädigen nicht nur die Tumorzellen, sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapierung von Krebs dar.
20
25
30

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-transkriptionale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20
25
30

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer 10 Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15 Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20 Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

25 Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das 30 Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).

20 Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

30 Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF,
5 M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:
10

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindearminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.
15

20 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

25 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.
30

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfundungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziertes Virus, Herpes Simplex Virus, Vaccinia Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfundungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

Die Zubereitungen der erfundungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

- 5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilesequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.
- 10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.
- 15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.
- 20 Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.
- 25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.
- 30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

- A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reporterexpression.
- Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.
- Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.
- Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilstücken unterschiedlichster Länge als Sonde.
- Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.
- Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

- A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reporteraktivität, um einen diagnostischen Wert zu erhalten;
- C. Vergleich des diagnostischen Werts mit Standardwerten für das Reporterkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;
- Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

- Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies
- A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Spraque Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).
- B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

10

15 Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur 20 14: nicht verdaut.

25

B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 10 5' GTTAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den 15 Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die 20 dargestellte Sequenz entspricht SEQ ID NO 2.

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25 Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymeschnittstellen sind wie folgt gekennzeichnet: S, SacI; X, Xhol. Die relative Anordnung der 30 fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

10

15

20

25

30

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promoters.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promoters in HEK-293 Zellen.

15 Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite sind die relative Luziferasaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

20

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpacket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in
10 Intron 2.

Tab. 1

3' Acceptor Sequence		5' Donor Sequence					
Intron	Exon	Exon	bp	Intron	Intron	Intr.	bp
No.	No.	No.	on	on	on	on	No.
5' flankierende Region							
caggcgctcccccgttag	GTTTCAGGCAAGCTGGGT	1	281	CGCCCCCTCTCCGCCAG	gtggggccctcccccggggtcg	1	104
catccatcttcgttttag	GTGTCCCTGCCCTGAAGAGC	2	1354	TGGCTGCGCAAGGCCAG	gtgaggagggtgggtggccgt	2	8616
gaggggctctatggtag	GGGTGGCTGTGTCGGGC	3	196	TGCAAAAGCATGGAATCAG	gtactgttatccccccgcga	3	2089
ACAGCACTTGAGAGGGTG	ACAGCACTTGAGAGGGTG	4	181	GTTCCGGAGAGAAGAGG	gtggccgtgtgttttgtttta	4	687
GGCGAGGGTCTCACCTCGA	GGCGAGGGTCTCACCTCGA	5	180	TGAGCTGTACTTTGTCAA	gtggggccgggggggggggg	5	494
GTGGATGTGACGGGGGGT	GTGGATGTGACGGGGGGT	6	156	CAAGGCCTTAAGGCCAC	gtaaagggttcaacgttgata	6	>4660
GTCTCTACCTTGACGACC	GTCTCTACCTTGACGACC	7	96	TGCGTCTGTCATCGAGCA	gtctgggcaactgcccgtcga	7	98U
AGCTCCTCCCTGATGAGG	AGCTCCTCCCTGATGAGG	8	86	CCGTGGCATCAGGGCAA	gtgaggtcagggtggccagg	8	248L
GTCCTACGTCCAGTCCAG	GTCCTACGTCCAGTCCAG	9	114	CGGGGATTGGGGGGACGG	gtgaggccctccctttccccc	9	1984
GCTGCTCCTGCGTTTGTG	GCTGCTCCTGCGTTTGTG	10	72	ACGCCAAAACCTTCCTAG	gtgaggcccccgtgtgtgtg	10	1871
GACCCTGGTCCAGGGTC	GACCCTGGTCCAGGGTC	11	189	TGGAGGGGACTACTCCAG	gtgagggcacccctggccggaa	11	380L
CTATGCCGGACCTTCATC	CTATGCCGGACCTTCATC	12	127	CTGTGTTCTGGATTTCAG	gtgaggcaggctgtgggttca	12	880
GTGAACAGCCCTCCAGCGG	GTGAACAGCCCTCCAGCGG	13	62	TCCCTGTCAGGGTACAG	gtgaggcccccaccaagggg	13	318L
GTTTCA CGCATGTGTGTG	GTTTCA CGCATGTGTGTG	14	125	CTGAAAGCCAAGAACCGAG	gtatgtcagggtggctggc	14	781
GGATGTCGCTGGGGGCCAA	GGATGTCGCTGGGGGCCAA	15	138	CTGGGTCACCTCAGGACAG	gcaagggtgggtggggggcc	15	536
CCCAAGAGCCAGCTGA GTCG	CCCAAGAGCCAGCTGA GTCG	16	664	TTTTTCAGTTTGTAAAAA	3'-flankierende Region		

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Spl	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
AP1	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

10

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

15
20
25

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschritte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

30

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml 10 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte 15 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.* 25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) 5 gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach 10 λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umkloniert (Beispiel 4).

Beispiel 3

- 15 Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).
20 Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

- 25 Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.
30

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-
5 Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

- 10 Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-
15 Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.
- 20 Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.
- 25 Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.
- 30

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklos P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10 Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-

15 ACTATAAGGGCACCGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20 kloniert.

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTTGAAGGCCTT-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genetischen hTC-Region identifiziert (s. Fig. 7).

Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelnsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

ACTTGAGCCC AAGAGTTCAA GGCTACGGT AGCCATGATT GCAACACCA CCGCCAGCCT TGGTGACAGA 70
 ATGAGACCT GTCTAAAAAA AAAAAGAAA ATTGAATA ATATAAAGCA TCTTCTCTGG CCACAGTGG 140
 ACAAAACAG AAATCAACAA CAAGAGGA ATTGAAAACT ATACAACAC ATGAAAATTA AACATATAAC 210
 TTCTGAATGA CGAGTGAGTC AATGAAGAAA TTAAAAGGA ATTGAAAATA TTATTTAG CAAATGATAA 280
 CGGAAACATA ACCTCTCAA ACCCACCGTA TACAGCAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA 350
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCGT TAATCCCAGC ACTTTGGAG 420
 GCCAAGGGCG GCAGATCGCC TGAGGTAGG AGTTCGAGAC CAGCCGTGACC AACACAGAG AACCTTGTG 490
 10 CTACTAAAAA TACAAAATTG GCTGGGCATG GTGGCACATG CCTGTAACTC CAGCTACTCG GGAGGCTGAG 560
 GCAGGATAAC CGCTTGAAACC CAGGGAGTGG AGGTTGCCGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC 630
 TGGTGAACAG GAGTCACCA CTGTCACAA AAAAAGAAA AAGTAGAGA AACTTAAAAT AACAACTAA 700
 GATGCACCTT AAGAAGACTAG AAAAGCAAGA GCAAACTAAA CCTAAATTG TAAAGAAA AGAAATAAA 770
 AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAT ACAAAAGATC AACAAATTAA AAAGTTGGTT 840
 15 TTTGAAAGAT ATAAAACAAA TTGACAAACCA TTGCCCCAGA CTAAGAAAAGGAGAAG ACCTAAATAA 910
 ATAAGTCAG AGATGAAAAG AGAGACATTA CAACTGTACAC CACAGAAATT AACAGGATCA CTAGAGGCTA 980
 CTATGAGCAA CTGTACACTA ATAAATTGAA AACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAAC 1050
 CTACCAAGAT TGAAACATGA AGAAATCCAA AGCCCAAAAC GACCAATAAC AATAATGGGA TAAAGGCCAT 1120
 20 AAAAAAAAGT CTCTTAGCAA AGAGAACGCC AGGACCCAACT CGCTCTCTGG CTGGATTTTA CCAATCATT 1190
 AAACAAAGAT GAATTCCAACT CCTACTCCTAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAAACTCAT 1260
 TCTACATGGC CAGTATTACG CTGATTCCAA AACCAGACAA AACACATCA AACAAACAAAC AACAAAAAAA 1330
 CAGAAAGAAA GAAAATACAA CGGCAATATC CCTGTGATGAAT ACTGTACACAA AATCTCTCAA CAAACACTA 1400
 GCAGAACAAA TTAAACAAAC CTTTCAAGG ATCATTGAT GTGATCAACT GGGATTATT CCAGGGATGG 1470
 AAGGATGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCACAA AAAATGAAAGT AACAAAACATA 1540
 25 TATGATTATT TCACTTTATG CAGAAAAGG ATTGTGATAAA ATTCTGACCC CTTCATGATA AAAACCCCTCA 1610
 AAAAACCCAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TCGGATCCCA GCACCTCTGG 1680
 AGGCCAAGGT GGGATGATTG CTGGCCCAAGG AGGAGTTGG ACTAGCTGCG GCAACAAAT GAGACCTGGT 1750
 CTACAAAAAA CTTTTTAAAG AATTTAGCCA GCGATGATGG CATATGCCGT TAGTCCCAGC TAGTCTGGAG 1820
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTGCGAGG TGCACTGAGC CATGAACATG TCACTGTACT 1890
 CCAGCCTAGA CAACAGAAC AGACCCACAT GAATAAGAAC AAGGAGAACAG AGAAGGGAGA AGGGAGGGAG 1960
 AAGGGAGGAG GAGGAGAAC AGGAGGTGGG GGAGAAGTGG AGGGGGAGG GGAGGGGGAGA GAGGAGAAC 2030
 AAGAAACATA TTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAACTGAA 2100
 AGCCCTTCTC CTAAGATCTG GAARATGACA AGGGCCCACT TTCAACTCTG TGATTCAACAA TAGTACTAGA 2170
 30 AGTCTCTGCT AGAGCAATCA GATAAGAGAA AGAAATAAA GGCATCCTAA CTGGAAAGGA AGAAAGTCAA 2240
 TTATCTGTT TGCAAGTGT ATGATCTTAT ATCTGGAAAAG GACTTAAGAC ACCACTAAA AACTATTAGA 2310
 GCTGAGCAG GGTACAGCAG GATACAAATG CAATGACAA AATATGAGT TATTCTATA TTCAACACG 2380
 AAACAATCTG AAAAAGAAC CAAACAAAGCA GCTACAAATAA AAATTTAACCA GCTAGGAATT AACAAAGAA 2450
 GTGAAAGATC TCTCAATGAA AAATCTTAAATGTTGATAA AAGAATTTGAGAGGGCACA AAAAAGAAA 2520
 40 AGATATTCCA TGTTCATAGA TTGGAAGAAT AAATACTGTT AAATGTCCTA TACTACCCAA AGCAATTTCAC 2590
 AAATTCATG CAACTCTCAT TAAATACTA ATGACGTTCT GTCACAAATG AGAAGAACAA ATTCATAGAT 2660
 TTGTACAGAA CCACAAAGA CCCAGAATAG CCAAGCTTAT CTCGACCAA AAGAACAAAAT CTGGAAGCAT 2730
 CACATTACCT GACTTCRAAT TATACACAA AGCTTACTA ACCCAACTA CATGGTACTG GCATAAAAC 2800
 AGATGAGACA TGGACCAAGA GAACAGAAAT GAGAATCCAG AACAAATTC ATGCACTCAT AGTGAACCTCA 2870
 TTTTGACAAAGG AGGTGCCAACG AACACTTCTT GGGAAAAAGA TAATCTCTTC AATAATGGT GCTGGAGGA 2940
 45 CTGGATATTC ATATGCAAA TAACAATACT AGAAACTCTG CTCTCACCCT ATACAAAAGC AAATCAAAT 3010
 GGATGAAAGG CTTAAATCTA AAACCTCAA CTTTCAACT ACTAAAGAGA AACACCGGAG AACACTCTCA 3080
 GGACATTGGG AGTCTCTGAG TAATCTCTGG CAGGACACAGG AACCCAAAGC AAAACACAC 3150
 AAATGGGATC ATATCAAGT AAAAGCTT TCGCCACCAA AGGAACAAAT CAACAAAGAG AGAGACAAAC 3220
 CCACAGAATG GGAGAAATATA TTGCAAACT ATTCTATCAA CAAGGAAATAA ATAACAGATA TATATAAGGA 3290
 GCTCAAACAA CTCTTAAAGA AAAACACCTAA ATAAGCTGAT TTCAAAATTG AAGCAAAAGA TCTGGTAGA 3360
 50 CATTCTCAA ATAAGCTCAT AACATGGCA AACAGGATC TCAACACCA CTGATCATCA 3430
 GAGAAATGCA AATCAAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAATGCTTTT TTCAAAAGAC 3500
 AGGAACAAAC AATAGGCCAT GAGGATGTTG ATAAAAGGAA ACCCTTGGAC ACTGTTGGT GGAATGGAAA 3570
 TTGTACACCA TATGGAGAAC AGTTGAAAG TTCTCTAAA AACTAAATTA AAGACTACCA TACAGCATC 3640
 CCATTGCTAG GTATATACCT CAAAAAGGG AATCAGTGT TCAACAGCT ATCTCCACTC CCACATTAC 3710
 TGCAAGCTG TTCATAGCAG CCAAGGTGG GAAGCAACCT CAGTGTCTCA CACAGACGA ATGGAAAAAG 3780
 55 AAAATGTGGT GCACATACAC AATGGAGTAC TACCGAGCCA TAAAGAGGA TGAGATCTG TCAGTTGCAA 3850
 CAGCATGGGG GGCACCTGGTC AGTATGTAA GTGAAATAAG CCAGGACAGG AACAGAACAC TTTTCATGTT 3920
 CTCCCTTACT TGTTGGAGCA AAAATTTAAAG CAATTGACAT AAGAATAGAG GAGAATGGT GTCTTAGAGG 3990
 60 GGTGGGGGAC AGGGTTGACTA GAGTCACAA TAATTATTG TATGTTTAAATTA AATTAACATAA AGACTATATAA 4060
 TGGTTGTTT GTRACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTTAC CCTGATGTGA 4130
 TTATACACCA TTGTATGCCAT GTATCAAAT ATCTCATGAT TGCTATAGA ATAACCCCTA CTATATTAA 4200
 AATTTAAATT TTAATGGCCA GGCACGGTGG CTCTGTCGG TAATCTCCAGC ACTTTGGAG GCGGAGGGCG 4270
 65 GTGGATCACC TGAGGTGAGG AGTTGAAAC CAGTCTGGCC ACCATGATGA AACCCCTGTCT CTACTAAAGA 4340
 TACAAAAATT AGCCAGGGCTG GGTGGCACAT ACCTGTAGTC CCAACTACTG AGGAGGCTGA GACAGGAGAA 4410
 TTGCTTGAC CTTGGGAGGC GAGGTTGCGAG TGAGCCGAGA TCATGCCACT GCACCTGAGC CTGGGTGACA 4480
 GAGCAAGACT CCATCTCAA ACAAACAA AAAAGAGAATTTAATTG TAATTTTAT GTACCGTATA 4550
 AATATATACCT CACTATATT AGAAGTTAAA ATTAAAACCA ATTATTTAAAG GTAAATTAAACCT AACTTAACTCA 4620
 70 AATTAAGAAC AATGTATGTT GGGTTTCTAG CTCTGAGGA AGTAAAGGT ATGCCACGA TGGCAGAAAT 4690
 GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAATGTGACTT AATTTAAC 4760
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAACG CCTAAACACCA CTGCTTAAATG TGGTGAAGG 4830
 TAATCTCAT TAATTACCAA TAATTACAGA TATCTCTAAATGAGCTGAGC AGAATGGCA CGTCTGATCA 4900
 CACCCCTCCTC TCATTCAACGG TGCTTTTTT CTTGTGTGCT TGGAGATTG CGATTGTGTG TTCTGTGTTG 4970
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAATGGT GTGATTTC CTCAGAAGAA TTAGAGTAC 5040
 75 TGGCAGGAAG CAGGTGGCTC TGTCGACCTG AGCCACTCTA ATCTCAAGG CTGCTCTGCC AAGACCCAGG 5110

	TGCAAGGCAG	AGGCCGTATG	ACCCGAGGAC	AGGAAGCTC	GGATGGGAAG	GGGGCATGAG	AAGCCTGCCT	5180
	CGTTGGTGA	CACCGATGA	AGTCCTTA	TTTACGCTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
5	AGGCGGCGAG	CGGGAATGCA	AGGAGTCAGA	AGCCCTCTGC	TCAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320
	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTC	AGGCACCTCG	AAAGTATGGCT	TAAATCTTTT	TTTCACCTGA	5390
	AGCAGTGA	AAAGGTATT	CTGAGGAAAG	CTTGAGTTAG	GTGCTTCTT	AAAAACAGAA	AGTCATGGAA	5460
	GCACCCCTCT	CAAGGGAAA	CCAGACGCC	GCTCTGCGGT	CATTACCTC	TTTCTCTCT	CCCTCTCTG	5530
	CCCTCGCGT	TTCTGATCGG	GACAGACTGA	CCCCCTGGA	GCTTACCTGA	CCCTCTCTG	5530	
	TGCAAGGGC	TCCACAGACC	CCCCCTGG	AGAGGAGACT	CTGAGCTGG	CTTAATAACA	AACTGGGATG	5600
	TGGCTGGGGG	CGGACACCGA	CGGGGGATT	CAAAGACTTA	ATTCATGAG	TAAATTCAC	CTTCCACAT	5670
10	CCGAATGGAT	TTGGATTTA	TCTTAATTT	TTCTTAATT	TCTCAAATA	ACATTCAAGG	CTGCAGAAAT	5740
	CCAAAGGCCT	AAAACAGAA	CTGAGCTATG	TTTGCAGGAG	TCTAACATT	TTTCAGAGGG	5810	
	ATTTTTCGCC	CTAAAGTACTT	TTTATTGGTT	TTCTAAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5880
	AGAGGGCTGG	GGGGCAGGGC	TATGAGCAGC	GCAGGCC	GGGGAGAGA	GTCCTCCGCC	TGGGAGGCTG	6020
15	ACAGCAGGAC	ACTGACCGT	CCTCCCTGG	AGCTGCACCA	TGGGGCAAC	CGAAGGGCGC	CACGCTGCCT	6090
	GTGACTCAGG	ACCCCATA	GGCTTCTGG	GCCCCCAC	ACTAACCCAG	GAAGTCACGG	AGCTCTGAAC	6160
	CCGTGGAAAC	GAACATGACC	CTTGCCTGC	TGCTTCTCG	GGTGATGGAA	GTGGTGTGCA	6230	
	GGAAATGGCC	ATGTAATTA	CACACTCTG	CTGATGGGG	CGCTTCTTC	CATCATTATT	CATCTCAC	6300
	CCCAAGGACT	GAATGATCC	AGCAACTTCT	TCGGGTGTGA	CAAGGATG	CAAAACTCA	TACAAAACACC	6370
20	ACTCTTATAC	TAGGCCCCA	GAGCAGCGG	CACACCCCTG	ATATATTAAG	AGTCACAGGAG	AGATGAGGCT	6440
	GCTTTTACGG	ACCCAGCTGG	GGTACACACA	CGGGCTGAC	AGTCTGTTTC	TCTAGACTAG	TAGACCTGG	6510
	CAGGGCACTCC	CCCAGATTCT	AGGGCTGTG	TGCTGCTTCC	CGAGGGGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGGTGC	CAACTGAGGC	CAGCCCTGC	TCCACACCC	CCGCTCCAG	GCCTCAGCTT	CTCCAGCAGC	6650
25	TCCTTAACCC	CTGGGTGGG	CGTCTTCCAG	CGTACTGTC	TCACCTGTC	CACTGTGTC	TGTCCTCAGGG	6720
	ACGTAGCTG	CACGGTTCT	CCTCACATGG	GGTGTCTGTC	TCTTCTCCCA	ACATCTCAC	GCCTGAGG	6790
	GAGGAGATTG	TGCGCTCCC	AGACTGGCTC	CTCTGAGGCT	GAACATGGCT	CGTGGCCCC	GATGAGGTT	6860
	CCTGGCGTCC	GGCTGCA	TGACCTTCT	TCCAGGCGC	TCCCCGCTC	CTGTGATCTG	CCGGGGCTG	6930
	CCGGTGTGTT	CTTCTGTT	TGTGCTCTT	TCCAGTCCA	GCTGCGTGTG	TCTCTGCCG	CTAGGGTCTC	7000
30	GGGGTTTTA	TAGGCATAGG	ACGGGGCG	GGTGGCCAG	GCGCTCTG	GGAATGCAA	CATTGGGTTG	7070
	TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCAGGCGA	CAGGGCTGG	GATGGAGCCC	CCGGCAGGG	7140
	CCCGCCCTTC	TCTGCCCCAG	ACCTTCTGC	CCCCCTCCCT	CTGGAACACA	GATGGCAGT	TTCCACAAGC	7210
	ACTAACGATC	CTCTTCCAA	AAAGACCCAG	ATGGCACCC	CTGGACATT	GCCCCACAGC	CCTGGGAATT	7280
	CACGTGACTA	CGCACATCAT	TGACACACT	CGCTGCGA	CGGACCCCC	CTGTTTATT	TTAATAGCTA	7350
	CAAAGCAGGG	AAATCCCTG	TAAAATGTC	TTAACAAAC	TGGTTAAACA	AAAGGGTCAA	TCCGCACGGT	7420
35	GGACAGTTC	TACAGTGA	GAGGAACATG	CCGTTTATAA	AGCCTGCAG	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAAC	TGCCCACCTC	ATGGGATACG	TACGCAACAT	GCTCCTAAAG	AAAGAATTTC	ACCCCATGGC	7560
	AGGGGGATG	TGAGGGGGT	TAAGGGGGT	TAAGGACGGT	GGGGGGCGA	GCTGGGGCT	ACTGCACCA	7630
	AAGGGAGTTT	CTGGTGTCTG	ATGGTATTGG	CTCAGTTATG	AGGACTAATC	CATAGGGGAG	TGGGGATGGG	7700
	GGAACCCGGA	GGCTGTGCA	TCTTGGCAT	GCCCCAGGT	CTTGGGAGC	ATAATGTC	AGAGATGCC	7770
40	ACGTCTGT	TCCCCAAC	CTGAGGACAG	AACCCGCCG	GCCCCAGGGC	CTTGCAGGT	GTGATCTCCG	7840
	TGAGGACCTT	GAGGTCTGG	ACTCTCGG	ACTACCTGCA	GCCCCAAA	GTAATCCAGG	GGTTCTGGG	7910
	AGAGGGGGC	AGGAGGGTCA	GAGGGGGCA	GGCTCAGGAC	GATGGAGGA	GTGAGCTG	GGCTGAAAGA	7980
	CGAGGGGG	CTTCGAGCC	AGGCGCTCAA	GGCCTCCAG	AAAGCTGAAA	AAGGGGGGAA	GGGACCCCTCC	8050
	ACGGAGCTG	CAGCAGGAG	GGCAGGCC	CCCTTAGCCC	ACAGGGCCC	ATCGTGGACC	TCCGGCTCTC	8120
45	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTTGTTGGG	ATTTGAGAA	GCAACAGGAA	8190
	ACCCATGCA	TGTAATCTA	GGGATTATTC	AAAACAAAGG	TTTACAGAAA	CATCCAAAGG	CAGGGCTGAA	8260
	GTGCTCCGG	GGCAAGGGAG	GGCAGGACAG	AGTGAATTTC	TTAGTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTATGCTCT	TGTTGCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400
	CGTCTCTGG	GTTCAGCAAA	TTCTCGTGC	TCAGGCTCCC	AAAGTACTG	GATTTCAGGC	GTGACCCAC	8470
50	ACACCCGGT	AATTTCAT	TTTATGAGA	GATGGGCTT	CACCATGTTG	TCTCAAGTGA	TCTCAAATC	8540
	CTGACCTCAG	GTGATGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACTGCACCT	8610
	GGCTTATTA	ACCTTCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGITC	8680	
	AATTTCCT	TCTACTCAGG	GTTACCTCC	TTTGATATT	TCTGAAATT	TTCGTTAGACT	GGGGATACAC	8750
	CGTCTCTG	CATATTCA	GTTTCTGTA	CCACCTGTTA	TCCCATGGG	CCACTCTGAG	GGGCAGCTGG	8820
55	GAGGCTGAG	GGCTGAGCTG	CCAGGGGGT	TCAGCATGTC	CACTGAGAAA	CTGATGAGA	ATCAGGGCCC	8890
	AAGTGTGGAC	ACTGTCCTG	ATCTCAATGT	CTCAGTGTG	GTCGAACAT	GTAGAAAATT	AAGTCCATCC	8960
	CTCTTACTCT	ACTGGGATG	AGCCCCCTTC	CTATCCCCC	CCAGGGGAG	AGGAGTTCT	CTCACTCTG	9030
	TGGAGGAAGG	AAATGATCTT	TGTGATTTT	CACTGCTGT	ACTGAATCCA	CTGTTAGGAA	TGTTAGAACA	9100
60	TTTGTGTTG	TTTGAGAGG	GGTTTACTC	TTGTTGCTCA	GCTGAGGGG	AGTGCATGG	CGCGATCTTG	9170
	GCTTACTGCA	GGCTCTGCC	CCCAGGTTCA	AGTGAATTCTC	CTGCTTCCG	CTCCCATTTG	GCTGGGATTA	9240
	CAGGCCCCC	CCACCATGAG	CAGCTTAATT	TTTGTTATT	TAAGTAGAG	GGGGGGGGGT	GGGGGTTACCC	9310
	ATGTTGGCC	GGCTGGTCTC	GAACCTCTGA	CCTCAGATGA	TCTACCTGCC	TCTGCTCTT	AAAGTGTG	9380
	GATTACAGG	GTGAGGCC	ATGCCAGCT	CAGAATTAC	TCTGTTAGA	AACTCTGGG	TCTGAGGAG	9450
	GAAGCTCACC	CCACTCAAGT	GTTGTTGGT	TTAAGGCCA	TDAGTAAATT	TTTTTATTGT	TGTTAGAACA	9520
65	CTCTTGTG	TTTACACTGT	GATGACTAAG	ACATCATGAG	CTTTCAAGG	ACACACTAAC	TGCAACCCATA	9590
	ATACTGGGGT	GTCTTCTGG	TATCAGCART	CTTCATGAA	TGCGGGAGG	CGTTTCTCG	CCATGCACAT	9660
	GGTGTAAATT	ACTCCAGCT	AATCTTCTG	TTCCATTCTT	TCTCTTAAATT	TGTGTTTTCT	9730	
	ATGTTGGCTT	CTCTGAGAG	AACAGCTGTA	AACTCTAAC	TAACTTTTGT	TGGAACAAAT	TTTCAAAC	9800
70	GCCCCCTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAA	AGGCTTAGGG	ATCACTAAGG	9870
	GGATTTCTAG	AAGAGGCCAC	TGTAATCTTA	AGTATTAC	AGACGAGGCT	AACCTCCAGC	GAGGCTGACA	9940
	CCCCAGGGAG	GGTGGAGGC	CTGTC	AAATTAAGG	AATTTCCTC	GGCAGTTCT	10010	
	AAAAGTAGGA	AAAGTTACAT	TAAAGGTTGC	GTTGTTAGC	ATTCAGTGT	TTGCGACCT	CGCTRCAGC	10080
	ATCCCTGCA	GGGCTCGGGA	GACCCAGAAG	TTTCTCGGCC	CCTTAGATCC	AAACTTGAGC	AACCCGGAGT	10150
75	CTGGATCTCT	GGGAAGTCT	CAGCTGCTCT	CGGGTGTGTC	CGGGGCCCCA	GGCTGAGGAG	GGACCACTGG	10220
	CCGTGTGGCT	TCTACTGCTG	GGCTGGAATG	CGGGCTCTCT	AGCTCTGCA	TCCGAGGCT	GGAGCCAGGT	10290
	GCCTGGACCC	CGAGGCTGCC	CTCCACCTG	TGCGGGGGGG	ATGTGACCA	ATGTTGGCCT	CATCTGCCAG	10360
	ACAGACTGCC	GGGGCCCCAG	GTCAAGGGCG	TTGTGCTGC	TGAGGAGGCG	CCGGTGGCGG	CCACAGCAGGA	10430
	GCCTGGCTGGCT	CCATTCTCCA	CCCTTCTCG	ACGGGACCGC	CCGGGTGGGT	GATTAACAGA	TTTGGGGTGG	10500

	TTTGCTCATG	GTTGGGGACCC	CTCGCCGCCT	GAGAACCTGC	AAAGAGAAAAT	GACGGGCCCTG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCTG	GCCCCGCCAG	GGAGCAATGC	10640
5	GTCTCTGGGT	TCGGTCCCCAG	CCGGCTCTA	GGCGCTCCGT	CCTCCCTCTT	ACGTCCGGCA	TTCGTGGTGC	10710
	CCGGAGCCCG	ACGGCCCCGG	TCCGGACCTG	GAGGCCAGCC	TGGGTCTCCG	GATCAGGCCA	GCGGCCAAG	10780
	GGTCGCCGCA	CGCACCTGT	CCCAGGGCC	CCACATCATG	GCCCCCTCC	CGGGTTACCC	CACAGCTTCA	10850
	GCGGATTCTGA	CCTCTCTCCG	CTGGGGGCC	CGCTGGCGTC	CCTGCACCC	GGGAGGCCGA	GCGGCCGCCG	10920
	GGCGGGGAGG	CGCGGCCAG	ACCCCCGGG	CGCGGCCAGG	CAGCTGGCTG	GTCGGGGCCA	GGCGGGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCAGG	ACCGCGCTC	CCACCTGGG	GAGGGACTGG	GGACCCGGGC	11060
10	ACCCGCTCTG	CCCCCTTCAC	TTCAGCTCC	GCTCCCTCG	GGCGGACCCC	GCCCCCTCCC	GACCCCTCCC	11130
	GGGTCCCGG	CCCAGCCCC	TCCGGGCC	CCCAGGCC	CCCCCTCTT	TCCGGGCC	CGCCCTCTCC	11200
	TCGCGGCCG	AGTTTCAGG	AGCGCTCG	CCTGCTGGC	ACGCTGGAA	CCCTGGCCC	GGCCACCCC	11270
	GCGATGCCG	CGCGCTCCCG	CTGGGAGG	GTGCTGCC	TGCTGCC	CAACTACCCG	GAGGTGCTGC	11340
	CGCTGGCAC	CTTCGCGG	CGCCCTGGG	CCAGGGCTG	GCGGCTGGT	CAGCGGGGG	ACCGGGCGGC	11410
	TTTCCGCGG	CTGGTGGCC	AGTGGCTGG	GTGCGTGC	TGGGGACGAC	GCCGCCCC	CGCCGCC	11480
15	TCCCTCCGG	AGGTGGGC	CCCCAGGCC	GGGGTCCCG	TGGGGTGA	GGCGGCCGG	GGGAACCCAGC	11550
	GACATGCCG	GAGCAGCG	GGCGACTCG	GGCGCTCC	CCCGAGGTG	TCTGCTGA	GGAGCTGGT	11620
	GCCCGAGTGC	TGCAAGGG	GTGCGAGC	GGGGCGAAGA	ACGCTGCC	CTTCGCTTC	GGCTGCTGG	11690
	ACGGGGGCTG	CGGGGGCCC	CCCGAGGCC	TCAACCAC	CGTGGCGACG	TACCTGCCA	ACAGGTGAC	11760
20	CGACGCACTG	GGGGGGAGG	GGGGCTGGG	GCTGCTG	CGCCGCTGG	GCGACGACG	GCTGGTTCAC	11830
	CTGCTGCCAC	GCTGGCGCT	CTTGTGCTG	TGGCTCCA	GCTGGCTTA	CAAGGTGTC	GGCCGCC	11900
	TGTAACAGCT	CGGCGCTGCC	ACTCAGGCC	GGGCCCCCCC	ACACGCTAG	GGACCCCGAA	GGCGTCTGG	11970
	ATCGGAACGG	CCCTGAAACC	ATAGCTCG	GGAGGCCGG	GGCGCTCC	GCTGCCAGC	CCCGGGTGC	12040
	AGGAGGCGC	GGGGCAGTGC	CAGCCGAGT	CTGCCGTTG	CCRAAGGGC	CAGGGCTGG	GCTGCCCTG	12110
25	AGGGGGAGG	GACGCGCTG	GGGGAGGCT	CCTGGGGCC	CCCGACCC	CGTGGTGGAC	TACCTGCCA	ACAGGTGAC
	TGTTTCTGT	CTGGTGTCA	TGCGCAGAC	GGCGAACAGA	GCCACCTT	ACGGGTGGAC	GCTCTCTGGC	12250
	ACCGCGCACT	CCCACCCATC	CTGGGGCCG	CAGCACACG	CAGGGCCCC	ATCCACATC	CGGCCACACC	12320
	GTCCCTGGG	CACGCCCTTG	CCCCCTTG	ACCGCGAGAC	CAAGCACTTC	CCTACTCTC	CAGGGACACAA	12390
	GGAGCAGCTG	CGGCCCTCT	TCTCTACTCG	CTCTCTGAGG	CCACGCTGA	CGGCGCTCG	GAGGCTCGT	12460
30	GAGACCATCT	TTCTGGGTT	CAGGGCTG	ATGCCAGGG	CTCCCGCG	GTTGGCCCG	CTGCCCGAG	12530
	GCTACTGGCA	ATAGGGGCC	CTGTTCTGG	AGCTGCTGG	GAACACCGG	CAGTGGCCCT	ACGGGGTGC	12600
	CCTCAAGACG	CTACTGGCG	TGCGAGCTG	GGTCACCCCA	CAAGCGGTG	TCTGTGCCG	GGAGAACCCC	12670
	CAGGGCTCTG	TGGCGGGCCC	CGAGGAGGAG	GACACAGAC	CCCCTGCG	GSTGCGCTG	CTCGGCCAGC	12740
	ACAGCAGCGC	CTGGCAGGTG	TACGGCTCC	GGGGCTCGG	CTCTGCCCC	CAGGCCCTG	CA:GGCCCTCTG	12810
35	GGGCTCAGG	CACAAACGAC	GGCGCTCT	CAGAACAC	AAGAAGTCA	TCTCCCTGGG	GAAGCATGCC	12880
	AAGCTCTCG	TGCAAGGAGT	GACGTGGA	ATGAGCGT	GGGACTGCC	TTGGCTGCC	AGGAGGCCAG	12950
	GTGAGGAGG	GGTGGCCGTC	GAGGCCAG	GCCCCAGAGC	TGAATGCA	GGGGCTCAG	AAAAGGGGGC	13020
	AGGAGAGGCC	CTGGTCTCC	TGCTCTCATC	GTACGTTG	TTTGGCTCA	GGACGTCAG	13090	
	TGGACACGGT	GATCTCTGC	TCTGCTCC	CTCTGTCCA	GTTTGTATA	ACTTACGAGG	TTCACCTTC	13160
40	CGTGTGATG	GACACGGGT	TCCAGGGCG	CGAGCAGGCA	GCACTGAA	GAGGAGGCTG	GGGGCGGCAG	13230
	TGGAGCCGG	TTGCGGCAA	TGGGGAGAAG	TGCTGTTGAG	CAACAGCTG	CTGGCGAGG	TGCTGCGAGG	13300
	TTACCTATAA	TCCTCTTCG	AATTCAAGG	GTGGGAATGA	GAGGTGGGG	CGAGAACCCC	CTCTCTCTGG	13370
	GGGTGGGAGG	TAAGGGTTT	GCAGGTGCA	GTGTTGAGG	AAATAGCAGG	TTTGTGTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTG	GGTGGCTC	CCAGGTAARTC	CCAGCACTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TCAGGAGTT	GAGACGAGC	TGACCAACAT	GTTGAAACCC	TATCTGACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGGTG	TGTGCTGC	TAATCCCG	TACTTGGAGG	GCTGGAGGAG	GGAATACACT	13650
	TGAACCCAGG	AGGCGAGG	TGCACTGAG	TGAGATTG	CCATTTGACT	CCAGCTGGG	CGACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTG	GCAGGACAG	GTAGAGGG	GGGAGATAAG	13790
	ACTGTTCTC	ACGACAGATO	CTGTCCTCC	CTTTAGGTG	GAAGAGGGG	ACATGGGAGG	AGGAGCACAGC	13860
50	AGATGGCTCC	ACCTCTGAG	GAAGGGACAG	TGTTTGTGG	TGTTGAGG	ATGGTCTGTC	TGGGCCCTGC	13930
	CGTGTCCCCA	CCCTGTTTT	CTGGATTG	TGTTGAGG	CCTCCGCTC	AGCCCCCTT	TGGCTCCCAG	14000
	TGCTCCCAAG	CCCTACCGG	GCAGCTGAGA	GAAGTCCCAG	TTTACCCAAA	CTCCCAAGAC	14070	
	ATGTAAGACT	TCCGGCCATG	CAGACAGG	GGGTGACCTT	CTTGGGCTC	TTTTTTCT	TTTTTTCTT	14140
55	TTATGGTGC	AAAAGTCATA	TAACATGAGA	TTGGCACTC	TAACACCGT	TTCTGTGAC	AGTGCAGAAT	14210
	TGCTAACTCG	GGGGTGTAA	CACCGAGGTG	CTTGAATG	TGCGCTCTTG	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCTCA	CACCTCTG	GGTCTGAGG	GACCCAGCGG	AGTCAGATAA	GGCTCATGCA	14350
	ACCCAGTTT	GCTTTTGTG	CTCCAGCTC	CTTCGTTGAG	GAGAGTTG	GTTCTGTAT	CAGGACTCTG	14420
	CCTGTCTATG	CTGTCTCTG	ACCTCAGATG	AGGTCAACAT	CTGGCCCTG	CTTATGCAAG	GAGTGGAGG	14490
60	TGGTCCCCGG	GTGTCTCTG	CAAGGTGAGG	GTGAGTGA	CGTGGCCCC	GTTGTCTCT	GTCACGTGTA	14560
	GGGTGAGTGA	GGCGCGGCC	CCGGGTGTC	CTGTCCTG	CAGCGTATT	GAGGTGTGG	CCCCGGGTGT	14630
	CCCTGTCACTG	TGTAGGTGA	GTAGGGCC	ATCCCGGTC	GTCCCTGTC	CGTGTAGG	GAGTGGAGG	14700
	TGGTCCCCGG	GTGTCTCTG	CCGGTCAAGG	GTGAGTGA	CACTTCTCCC	GTCACGTGCA	14770	
	GGGTGAGTGA	GGCGCGGTG	CCGGGTGTC	CTCTCAGGTG	TAGGGTGA	GAGGCGCGGC	CCCAGGGTGT	14840
	CCCTGTCACTG	TGTAGGTGA	GTAGGGCC	GTCTCTCTG	GTTGTGAGG	GAGGCGCGGC	CCCAGGGTGT	14910
65	CTGTCCCCGG	GTGTCTCTG	CAAGGTGAGG	GTGAGTGA	CGTGGCCCC	GTTGTCTCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCGTGTG	CTGGGTGTC	CTGTCCTG	TAGGGTGA	GAGGCTCTG	CCCCAGGGTGT	15050
	CCTTGGCTG	TGCTCACTG	AGCTGCTC	TGAATGTTG	CTCTTCTAT	AGCCACAGCT	GGCCCGGGTTG	15120
	CCCCCTGCT	GGGTGAGTGA	TGCAAGGCA	GTGCTGTC	CCAAAGCTT	TTTTCTGAT	GTCGGCTCT	15190
	TCTTGGTCA	CTCTCCGTC	CATTTGCTA	CGGGGACACG	GGACTGCAAG	CTTCGCTC	CCCGGTGCCA	15260
70	GGCAGCTGAG	CCACAGCTT	AGGTCCCGT	GCTCTCTG	GCCCCCTG	GTCACCCACAG	TGCCCCGCCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCTCAGT	TGTCTCATG	CGAGGCTGGA	CTCTGGCTG	CCTGTGCTG	15400
	CTGCCACGTG	TGCTGAGG	CATCCCGAA	AGGGTTCT	GTGCCCTGAA	GGAAAGCAAG	TCACCCCGAC	15470
	CCCCCTCACT	GTCCCTGTTT	CTCCCAAGCT	GCCCCCTCTG	TTGGGCCCC	TGGGTGGGTG	GCAACGCTTG	15540
	TCACCTTATT	CTGGGACCTT	GGCGCTATT	GCTTAGGTG	GGCTCTGCT	CCAGTCCCC	CCTCACATGG	15610
	ATTGACGTC	AGCCACAGGT	TGGAGTGTCT	CTGTCCTG	CTGTCCTG	GACCCACGTG	GAGGGCCCGT	15680
	GTCTCCGCA	GGCTTCTGCA	GACTTCCCTC	TTGGGTCTTA	TTTTTGAAT	TCACTGATTT	ACCTCTGACG	15750
75	TTTCTATCTC	TCCATTTGAT	GCTTTTCTT	GGTTTATCT	TTCACTTCTT	TTCTAGCTTC	TTAGTTAGT	15820
	CATGCTTTC	CCTCTAAAGT	CTGCTTAC	TGCACTCTG	TTTTTGATGT	GAAGTAATCT	CAACATCAGC	15890

CACTTTCAAG TGTCTTAAAT AACTTCAAAT GTGTTAACAT TTCTTTAAAG TATTCTTATT CTGTGATT 15960
 TTCTTGTG CACGCTGTG TTTGACGTGA AATCATTTC ATATCAGTGA CTTTTAAGTA TTCTTTAGCT 16030
 TATTCTGTGA TTCTTGTGAG CAGTGAAGTTA TTGAAACACT GTTTATGTC AAGATATGTA GAGTATCAAG 16100
 ATACGAGAG TATTCTTAAAGT TATCATTATA TTGATGACTT CTAACCTAGT TGTGATGTTG TCTGTATAAT 16170
 5 ACCAATTATTG TGAAGTTTG GGAGCCTTGC TTGTTGATCT AGTGTGTCA TGTTTCCAG AACTGTCCAT 16240
 TGTAATTG ACATCCTGTC AATAGTGGGC ATGCATGTC ACTATATCCA GCTTATTAAAG GTCCAGTGC 16310
 AAAGCTCTGT CTCCCTCTAG ATGCATGAAAT TTCCAAAGAAC GAGGCCATAG TCCCTCACCT GGGGGATGGG 16380
 TCTGTTCAT TCTCTCGTGT GGTTAGCATT TGTGAGGCG ATTGTTAGCT GCATGCACCGT GGTAGAATT 16450
 TTATCTTCT GATGAGTGAAT CTTTGGAG ACTTCTATGT CTCTAGTAAT CTAGTAATTCTTTTTAAA 16520
 10 TTGCTCTTATG TACTGCCAAT CTGGCTTCTT TTGAGTATTG ATTTCTCTG TGTTGCTGTT TTCTGCCCC 16590
 AATTATATATA TATATATATA TTTTTTTTTT TTGAGACAGA GAGTCTGGT CTGTCGCCCA GGGTGACTGC 16660
 AGTGGTGTGA TCACAGGTCA GTGTAACCTT TACCTTCTGG CCTGAGCGGT CCTCTCACCT CAGCCTCTG 16730
 AGTAGCTGGA ACTGCAGACA CGCAGGCTA CACCTCTCA ATTTTAAAT TTGTTCTGGA GACAGGGCT 16800
 TGCTGTTG CCGAGGCTGG TCTCAAACTC TTGGAACCAAAGGATCACAT TACCTGGCT TCCCAAATG 16870
 15 CTGAATTACA GGCATGAGCC ACCATGTCTG GCTTAATTTC CAAACACTTTT ATATCTTAT AGTGTGGGT 16940
 TGTCCTGTA ACAGCATGTA GTGTAATTG CAATCCTAGTC TGACAGTGTG TGTTTAACTG GATAACCTGA 17010
 TTATTTCTA TTTTTTTTTT TTGAGACAGA GAGTCTGGT CTGTCGCCCA GGGTGACTGC 16660
 CCTCGTTCCC TTGTTCTCA CCACCTCTGG GTGTGCCATG TGCGTTTCTC GCGGAGTGTG TGTTGATCCT 17080
 20 CTGCTGCTC CTCGTCCTAG GGGCATTGG TTTTATTCTC TTGCTGCTGTT TGTTACCCCC TGATCTTTT 17220
 ATTTGCTGTTG TTGTTTGTG TTTATTGAGA CAGTCTCACT CTGTCACCAAGGCTGGAGTG TAATGGCACA 17290
 ATTCGGCTC ACTGCRACCT CTGCTCCCTC GGTTCAAGCA GTCTCATTC CTCAACCTCA TGAGTAGCTG 17360
 GGATTCAGCG CGCCACCCACG CACGGCTGGC TAATTTAGTG ATTTTTAGTG GAGATAGGGT TTACCATGTT 17430
 TGCCGAGCT GGTCTAACAC TCTGACCTC AAGTGTATCG CCGCCCTGGT CCTCCACAG TGCTGGGATT 17500
 25 ACAGGTGCAA GCCACCGTGC CGGGCATACC TTGATCTTTT AAAATGAAGT CTGAAACATT GCTACCCCTG 17570
 TCTCTGACCA TAAGACCCCT AGTGTATTGTT AGCTGGCAGC ACCCCCCAGC CTGTCGGCTG TTTTCCCTGC 17640
 TGACTTAGTT CTATCTCAGG CATCTTGACA CCCCCACAAAG CTAAGCATTA TTAATATTGTT TTCCGTGTT 17710
 GAGTGTCTG GTAGCTTGC CCCGCCCTG CTTTCTCC TTTGTTCCCCT GTCTGCTTC TGTCAGGC 17780
 CCCGCCCTG GGGTCCCCCT CTTGTCCTC TGCGGTTGTT TTGTCCTGTT TTGTCCTGG TAACACCCAG 17850
 30 CTTTACCTG TCTGGCTCCCT ATGGCATCTA GGCACGTCGG CGGACCTCTGG CTTATGATGC ACAGATGAAG 17920
 ATGTTGGAGAC TCACGAGGAG GGCCTGTCATC TTGGCCCTGG AGTGTCTGGA GCACCCACGTG GCCAGCGTC 17990
 CTAGGCTGAGT GAGTGCAGC AACGTCGGCT CGGGCTGGT TGACCTGCAAAACCCAGG TGTCGGGG 18060
 TCTGTTGCTC CGGGCTGTC GAGTTGAAAAA CTCGGCAAAAC CTGGCGTGTG CGCGCAGCTG TGACGGTGT 18130
 GCGCTGGCGGG GGAGTGTCTG CTTCTCCCT TCTGCTGGG ACCAGGACA AAGGATGAGG CTCCGAGCCG 18200
 TTGTCGCCA ACAGGAGCAT GAGCTGAGCC ATGTTGATAAAT TTGTTAAATG TCTAGGCTGG CGCCGGTGGC 18270
 35 TCACGGCTGT AATCCCGAAC CTTTGGAGG CCAAGGCGGG TGGATCAGGA GGTCAAGGGG TCGAGACCAT 18340
 CCTGGCAAC ATGATGAAAC CCCATCTGA CTAAAACAC AAAAATTAGC TGGGGCTGGT GGGGGTGGC 18410
 TGTAATCCCA GCTACTCGG AGGCTGGG AGGAGAATG CTGAAACCTG GGAGTGGAA GTGCACTGA 18480
 GCCGACATTCG CACCACTGCA CTCCAGCTG CCAACACAGC GAGACTCTGT CTCAAAAAAAA AAAA 18550
 40 AAAAAAAA AATTCTAGTA GCCACATTAA AAAAGTARAA AAGARAAGGT GAAATTATG TAATAATAGA 18620
 TTGTTACTGAA GCCCACGTC TCCACACCTC ATCATTTAG TGCTTATAGG TGTTGAGCTG CACTTCACAGG 18690
 ACATTTGACA TTTTTTGAGC TTGTCCTGCG GGATCCCTGGT TGAGCTGCTC TGTCGTTGGC ATTCGCCCC 18760
 GGACTCTGTC GGCTTCCCAT GGCCATGGCT GTGTACCAAG ATGGTGCAGG TCCGGGATGA GTCGCCAGG 18830
 CCCTCAGTGA GCTGGATGTC CAGTGGATGTC ATGGTGCAGG TCCGGGATGA GTGGCTGTG CCGCTGCTGG 18900
 AGCTGGATGT GTGGTGTCTG GATGGTGCAGG TGAGCTTCAAGGCTGGT GCGCTGGTGG AGCTGGAGGT 18970
 45 ATGGAGTGGC GATGATGTCAG GTCCGGGGTG AGGTGCGAG GCGCTGCTG GAGCTGGATG TGTTGTTCT 19040
 GGATGGTGA GGTGAGGGGG GAGGTCTCAAGGCTGGT GAGCTGGAGG TATGGAGTCC GGATGATGCA 19110
 GGTCGGGGGT GAGGTGCCA GGGCCCTGCTG TGAGCTGGAT TGTTGTTGTC TGAGTGGTGC AGGTCTGGG 19180
 TGAGGTCACTC AGGGCCCTGGC GTGAGCTGGG TGTCGGGTGT CTGGATGGT CAGGCTGGA GTGAGGTGGC 19250
 CAGACGGTGC CAGACCATGC GGTGAGCTGG ATATGCGGT TCCGGGATGA GCGAGTCTGG GGTGAGGTGG 19320
 50 CCAGGGCTG CTGTGAGTC GATGTTGGGGT CTGGGGATGC TGCAAGTCGG TGTTGAGGTC ACCAGGCCCT 19390
 GCTGTGAGCT CGATGATGTC TGTCGAGATG TGTCAGGCTC GGGGTGAAGG TGCCCGAGCC CCTGCTGGT 19460
 AGCTGGATGT GTGTCGTCG GATGGTGCAG TGCTGGAGT AGGTGCGAG GCGCTGGTGC AGCTGGAGTGT 19530
 GCACTGTCGA GATGGTGCAG TGCCGGGGTG AGGTGCGAGG ACCTCTGGGT GAGCTGGATG TGCGGTGTT 19600
 55 GGATGGTCA GGTCTGGAGT GAGGTGCCA GGGCCCTGGT GAGCTGGATG TATGGAGTCC GGATGGTCC 19670
 GGTCGGGGGT GAGGTGCCA GACCCCTGTC TGAGCTGGAT TGCGGGATGC TGAGTGGTAC AGGTCTGGAG 19740
 TGAGGTGCCG ACACCCCTGCT GTAGGGCTGT CGGGATGGTG CCGGGATGGT CAGGTCAGGG GTGAGGTCTC 19810
 CAGGCCCTG TGAGGTGCCG GGTATGGAGT CGGGATGATG CAGGTCGGGG TGAGGTCGGC CAGGCCCTGC 19880
 TGTAACCTG ATGTCGGCG TCTGGATGTC GCAAGTCGG TGTTGTTGTC CGAGGCCCTC GGTCAGCTGG 19950
 60 AGGTATGGAG CTCGGATGATG CGCAAGTCGGG GGTAGGGCTC CGAGGCCCTG TGTCAGCTG TGAGTGGGG 20020
 GTCTGGATGTC TGCAAGTCGG GGGTGGTGTG CGCAAGGCCCT CGGTGAGCTG GAGGTATGGA GTGCGGATGA 20090
 TGCAAGTCGG GGGTGGGTT CGCAAGGCCCT GCTGTGAGCT GGTATGTCGTT TATCGGGATG TGTCAGTCGG 20160
 GGGTGGAGTC CGCAAGGCCCT GCTGTGAGCT GGTATGTCGTT TATCGGGATG TGTCAGTCGG 20230
 CACCAAGGGC TCGGGTGAGG TGTTGTCGG TGTCAGGCTG GGTGAGCTG CCGGGTGAAGT TGCCAGGCC 20300
 CTCGGTGAGC TGGATGTCG TGTCGGGGCTG GTCCGGATGTC TGCAAGTCGG GGGTGGAGTC GCTAGGCCCT 20370
 65 TGTTGGCTG GATGTGGCTG GTCCGGATGTC TGCAAGTCGG GGGTGGAGTC CGCAAGGCCCT TGTTGAGCTG 20440
 GATGTGGCTG GTCTGATGTC TGCAAGTCGG GGGTGGAGTC CGCAAGGCCCT TGTTGGGGCTG GATGTGGTGT 20510
 GTCCGGATGG TCGAGGTGCCG CGGTGAGGTC GCGCAAGGCCCT GCTGTGAGCT GGTATGTCGG TGTCAGTCGG 20580
 GTGCAAGTCGG GGGTGGAGGT AGGCAAGGCCCT TTGGGTGAGC TGAGTGTGGG TGTCAGTCGG GGTGAGGTC 20650
 CGGGGTGAGG TCGGGCAGGCC CTGGGGTAGTG CGGTGATATGC GGTGTCGGGA TGTTGAGCTG CCGGGGTGAG 20720
 70 GTCACCAAGGC CCTGGGGTTA GCTGGATGTC CGGTGTCGG TGTTGTCAGG TCCGGGGTGA GGTGCCCCAGG 20790
 CCCTGCTGTG AGCTGGATGT GCTGTATCG GATGTGGCAG TGCCGGGGTG AGGTGCGAG GCGCTGCACT 20860
 GAGCTGGATG TGCTGATTC GGTGAGTCGA GGTCTGGCGT GAGGTGCCA GGGCCCTGCGG TTAGCTGGAT 20930
 ATGCGGTGTC GGATGGTCA GGTGGGGGT GAGGTCAACCA GGGCCCTGCGG TTAGCTGGAT GTGCGGTTGTC 21000
 CGGATGGTGC AGGTCTGGG TGAGGTGCCG AGGCCCTGCT GTGAGCTGGA TGTCAGTCGG TCGAGTCGG 21070
 CAGGTCGGGG GTGAGGTGCCG CAGGCCCTGC GGTGAGCTGG ATGTGCTGTA TCCGGATGGT GCAGGTCGG 21140
 CGTGAGGTGC CGAGGCCCTG CGGTGAGCTG GATGTGCACT GTACGGATGG TGCAAGTCGG GGGTGAGGT 21210
 75 CGCAAGGCCCT CGGGTGGGCT GTATGTGTG TGTCAGTCGG GGGGGTGAAGT CGCCAGGCC 21280

	TGGCGGTGAGC	TGGATGTTGTC	GTGTCCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TGCCCAAGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCCTGGT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
5	GATGTGGCGT	GTCCGGATGG	TGCAAGGCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAACCTG	GATGTGGCGT	21490
	GTCCGGATGG	TGCAAGGCTG	GGGTGAGGTC	ACGACCGCT	CCTGATCTG	GATGTGGCAT	GTCCCTCTCG	21560
	TTAAGGGGT	TGGCTGTGTT	CCGGCCGCG	AGCACCGCT	GGGTGAGGAG	ATCTCTGCCA	AGTCTCTGCA	21630
10	CTGGCTGATG	AGTGTGTACG	TGTCGAGCT	GTCAGGTCT	TTCCTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAACAGGC	TCTTTTCTA	CCGGAAGAGT	GTCTGGACCA	AGTTGCAAAG	CATTGGAATC	AGGTACTGTA	21770
	TCCCCACGCC	AGGCCCTCTGC	TTCTCGAAGT	CCTGGAACAC	CAGCCGGCC	TCAGCATGCC	CCTGCTCTCA	21840
	CTTGCCCTGTG	CTTCCCTGGC	TGTCGAGCT	TGGGGCTGGG	GCCAGGGGCC	CCCTCACAGG	CCTGGTCCAA	21910
15	GTGATTCTG	TGCAAGGCTC	TGACTGCTG	GAGCTCACGT	TCTCTTAAC	GTAAATCAG	GAGTTTGTGC	21980
	CAAGTGGCT	CTAGGGTTG	TAAAGCAGAA	GGGATTAAA	TTAGATGGA	ACACTACAC	TAGCTCTT	22050
	GGCCTTCCCCT	GGGATGTGGC	TCTCTCTTC	TGGGGCTGG	TTCCTCTT	TTTGGAGATGG	AGTCTCTACTC	22120
	TGTTGCCAG	GTCTGGAGTC	AGTGGCTAA	TCTGGCTCA	CTGCAACCTC	CACCTCTGG	TTTAAGCGA	22190
20	TTCACCAGGC	TCAGCCCTCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCCCTGCT	AATTTCCTGA	22260
	CTTTAGGAG	TCACCATG	GGCAGGCTG	GTCTGCAACT	CATGACCTCA	GGTGATCCAC	22330	
	CCACCCCTGGC	CTTCCAAAGT	GCTGGGTTA	CAGCTAACG	CACCTGCC	AGCCCCCGAT	TCTCTTTAA	22400
	TTCATGCTGT	TCTGTATGAA	TCTTCATCT	ATTGGATTA	GGTCATGAGA	GGATAAAATC	CCACCCACT	22470
25	GGCGACTCAC	TGCAAGGAGG	ACCTGTGCA	GGGACACCTG	GGGATAGGAG	AGTCCACCA	TGACCTTAAC	22540
	TCTAGGTGGC	TGCAATTGAA	TGGCTGTAG	TTTTGCTG	CAATTCCTGG	CTGTAGAGAG	TGTGAGATTG	22610
	TGACAGATTC	AAGCTGGATT	TGCACTAGT	AGGGACGGG	GGCCTGGTCT	GGGAGATGCC	AGCCTGGCTG	22680
	AGGCCAGGAC	ATGGTATTAG	CTTCTCTGGT	TGGGGCTGG	GCTGAGTGT	GGGGGCTTTA	GTCAAGAAGAT	22750
	CAGGGCTTC	CCAGCTCC	TGACACTCG	AGTCCCTGG	GGGGCTTGTG	ACACCCCATG	CCCCAAATCA	22820
30	GGATGTCTG	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAAACACAGC	CTCTGGCTG	GGGACCCCGA	22890
	CGTGTGCTGC	GGGCAATTTC	CTTCTGCTG	GGGGAGGGCT	AGGGGCTTCC	CTGTTGGGAAAC	AAGTTAAATAC	22960
	ACAATGCAAC	TCTACTAGAC	TGACAGCTG	TTTAATGCTG	TGGCAACCAA	CATGGTATT	TGACCAAGTAT	23030
	TTTGGAAAGA	ATTTATTGG	GGTACCGGA	AGGAGCAGAC	AGACGTGGT	GTCCCCAAGA	TGCTCTTGT	23100
	CACTACTGG	ACTGTGTTG	TGCGCTGGG	GGCTTGGG	CCCCCTCTCC	CTGGACAGGG	TACCGTGCCT	23170
35	TTTCTACTCT	GTCTGGCTG	CCGGCTCGG	TGACGGCACC	AGCTCGGAG	CACCCGGGC	CCCACTGTC	23240
	ACGGAGTGC	AGGCTCTCAG	CCACAGATGC	CCAGGTCAG	GTGTCGGC	TCCAGCCCC	GTGCCCCCAT	23310
	GGGTGGTTT	GGGGGAAAGA	GCAACAGGCA	GGGGCTGCA	GAGACTGTG	GGCTCATGAG	AGGTGATTCT	23380
	GCTCTTCTGC	TGAGCTGCC	TGACGAGCT	CTCCGGCCCT	CTTCTATGCA	AGGGATGTG	CTCTTCTAC	23450
40	CTGGGGTTC	TGCCCTGGG	CAGCCTTGGG	CTACCCCGAT	GGCTGTACCA	AGGGACAGG	CATCCGTGT	23520
	GGAGGGGGAT	GGGTTACGCT	GGGGCATGG	GCACGCTCTC	GGGGGCTTCC	CTGGACGCTA	TGTTGGGACA	23590
	GTCACTCTGG	GGGTTGACGG	CCGGACTGG	TCGCCCCAGG	GGTGTATATA	GGACCAAGTG	TCCAGGTGCC	23660
45	CTGAAGTAG	AGGGGCTCTC	AGAGGCCT	GGCTGGCATG	GGTGGACGTG	GGCCCGGGCA	TGGGCTTCAG	23730
	CGTGTGCTGC	CGTGGGTGCG	CTAGGGCTC	ACTGAGTGTG	TGGGGCTTCC	TGACCTTC	23800	
	CCTAGTGTGT	TGTCCTGCTG	AGAACGCTC	CTGAGGGCT	CTTATTCGCA	GAACAGCACT	GAAGAGGTG	23870
	CAGCTGCGG	AGCTGTGCGA	AGCAGAGTC	GGGAGCATC	GGGAAGCCAG	GCCCCCGCTG	CTGACGTCCA	23940
50	GACTCCAGT	CATCCCCAAAG	CTCTGACGGG	TGTCAGTCTG	TGTCAGTGT	GACTACGTCG	TGGGAGCCAG	24010
	AACCTTCCC	AGAGAAAAGA	GGGTTGGTGT	GCTTTGGTT	AACTCTTT	TTAACACAGA	GTGCTTGTG	24080
	CCCCCACATT	TGTTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCAGG	GGACACAGC	AGGGCCATGG	24150
	CACGGCGGCCA	ACCCATTGTC	GGCCACAGTC	GGGTCGGGCA	GGTGGCTG	CTCCCAAGAA	AGCAGCGTGG	24220
55	GGGTGTAGGG	GGGAGCTCTG	GGGAGGGAC	AGGCTCTGAG	GACCAAGAAG	AGCAGCCGGG	CCAGGGCTG	24290
	GATGAGCAC	GGCCCGAGGT	CCTGGATCCG	TGTCCTGCTG	TGGTGCAG	CCTCCGTGCG	CTTCGCTTA	24360
	CGGCGCCCGG	GGACCCAGGG	AGCAGCTGCA	GGGACTGGG	GGGGCTGAG	GATCCCTGAG	CTTGGCCAC	24430
	GGGCTCTGCA	CCCCACCCCT	GTGCTGCGG	TGGCTGGCTG	GACCCCGTGA	TCTGAGGAGA	GTGTTGGGTG	24500
60	AGGTGGAGCAG	AGGTGTGCGA	TGAGGATCCC	GTGTCACAGA	CACATGCGC	CAGGAACCCG	TTTCAACAG	24570
	GGTGTGAGGA	AGCTGGGAGG	GGTTCTAGGT	GGGGGGCTG	GGTGGCTG	GGACTCTGGG	AGGGGGCTGCT	24640
	TCTCCCCCTGG	TGCCCTATGG	TGGGGTGGG	ACTTGGGG	ATCCACTTC	CTGACTGTCT	CCCATGCTGT	24710
65	CCCCGCCAGG	CGGAGCTCT	CACCTCGAGG	GTGAGGGAC	TGTCAGGGT	GCTCAACTAC	GAGGGGGGCC	24780
	GGGGCCCGGG	CTTCTGGG	GGCTCTGTC	TGGGGCTGGA	CGATACCTAC	AGGGCTTCCG	GCACCTTCTG	24850
	GCTCGGTG	GGGGCCCGGG	TGAGCTGTC	TGGTCAAGG	TGGGGCTGGG	GGACCCCGT	GTGTTGGGTG	24920
	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GGCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCT	24990
70	GGGTGGGCCG	CAGGGAGTGC	AGGTGACCTC	GTCACTGTG	AGGACACACC	TGGCACCTAG	TGGGGAGGCC	25060
	TTCAGCTTCTT	TCTCGACAC	ATGGGGCGGA	CTGTCACCC	TGACTGTCTG	GGCTCTTATT	CCAAGAGGG	25130
	GTCCCCACTGG	ATTCAGTTT	CCGTCAGAGA	AGGAACCCCA	ACGGCTCAGC	CACCAAGCC	GGTGCCTTG	25200
	CACCCCACTG	CTGAGGCAAG	GGTCTCTGT	CTGAGGGCTC	GGGGCTGAG	CACAGCCGCC	CCTGGCCCTTG	25270
	GGGTGTGAG	GGGGGGGGT	GGGGGGAGA	GGGGGGAGA	AGAGAGAGA	AGAGGGGG	CACAGCCGCC	25340
75	ATGTCTGAGT	TCTGGGTGG	CCACTGTCAG	TCTCTCTCC	TCCACTCACA	CAGGTGGATG	TGACGGGCC	25410
	GTACGACACC	ATCCCCAGG	ACAGGCTCAC	GGAGGTCTAC	GGCAGCATC	TCAACACCCA	GAACACGCTAC	25480
	TGGCTGCGT	GGTATGCCGT	GGTGTGCTG	GGTGTGCTG	GGTGTGCTG	CTATGGCATG	CTACGGCTCT	25550
	TAAGGTTCA	GTGTGATAGT	CGTGTGCTGG	ATGTGTGTC	CTGGGATATG	AAATGTGCTA	GAATGCACTC	25620
	GTGTCTGTA	TGGCTGCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	TGTGTTGGCA	25690
	GGTGTGTC	TGGGTGCTG	TATCTGTC	GGTCATATT	TTGGTGTGTC	TGTGTTGTC	ACGTGTGTG	25760
	CCATGGTGT	TGTGGCTG	GTGTCATGT	GTGTTGTC	GTGACAGCTG	CATGTTCATG	CTGTTGCTG	25830
	CATGTCTGT	ATGTGCTGAT	TTGTGCTG	TTGTGCTG	TTGTGCTG	ATATGGCTGT	CTATGGCATG	25900
	GGTGTGTC	GGCCCTTGGG	CTTACTCTT	CCTCTCTCAG	GCATGGTCC	CACCATGTC	CTACGGCTCT	25970
	CGGGTGTCTGG	TTGGGGAGG	TCCACATCA	GGGTCTCAG	TTCTAGCATG	GGTGCCTCTG	TCCCTGTCACA	26040
	GGGCTGGGCC	TTGGAGACTG	TAAGCCAGT	TTGAGGGAGG	AGTAGGGATG	CTGGTGGTAC	CTTCTGGAC	26110
	CCCTGGCACC	CCCAGGACCC	CAGTCTGCC	TATGCCGCT	CCATGAGATA	TAGGAAGGCT	GATTAGGCC	26180
	TCGGCTCCCCG	GGACACACTC	CTCCCCAGAC	GGCCGGGGGG	CTTGGGGCTC	GGCAGGGGG	AAAGGGGCC	26250
	TGGGCTTGGG	TTCCCACCA	GTGGTCA	GCACGCTGGA	GGGTAAGCC	CTCRAAGTC	TGCCAGGCC	26320
	GGGTGCAAGAG	GTGAAAGAAT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCC	CATGTGAAA	CCCAAGAGGA	26390
	CCTCTTCTC	TGACTTCTTG	AGCT					26414

Contig 2:

5	TGTGGGATTG	GTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGTAACAC	70
	AGAGTCAAG	GCGAGTTTC	TTCCCTGAGT	GGGTCTGCAG	GTGCTCCAAC	AGCTTATTG	AGGAGACCAT	140
	ATCTTCTTT	GAACATGGT	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCTGG	GCTCCCTGTT	210
	CTGTTCTTC	CACTCTGGG	TCGTGTGGTG	CTCGCTGTG	TGTGTGGCG	GTGGCAGGG	CTTCAGGCC	280
	TCCTTGTGTT	CATTGGCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCCTTGG	AATTCCCTG	CGAGTGGAG	350
	GCTTTCTTC	TTCTTCTTT	TTTTTTTTT	TGATAACAGA	GTCTCGCT	TTTTGCCA	420	
10	GGCTGGAGTG	GTTTGGCTG	ATCTTGGCTC	ACTGCACCT	GTGCTCCCTG	AGTCAGGAA	ATTCTCTGC	490
	CTCAGCTCC	CAAGTAGCT	GAATTATAGG	CGCCACAC	CATGTCAC	AATTTTGTA	ATTTTAGTAG	560
	AGACGAGGT	CTTCCATGT	GGCCAGGTG	GTCTCGAAC	CCTGACCTCA	GGTGTACCTC	CCACCTGGC	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CCGCCGCC	CGGCCGAGAC	TCGCTTCC	CAGCTTCCG	700
	GAGATCTGA	CGCATAGCT	CTCGACGCT	TGGTGTGAC	AACCTCCGTT	TTCCCTCTCC	AGGTCTCGCT	770
15	AGGGGCTTT	CCATTCATG	ACTCTCTCA	CAAGAACGT	TCACGTGTC	TGATTTCCG	GCTGTTCT	840
	GGCTTAATTG	TGTCGTGTT	TTATCGATGG	CTCCCTTCA	TTCTCTTAA	GCTTGTGTTA	TTGTGTTT	910
	TCCGGCTCT	TGAAGAAA	GTTGATGTTA	AACTTCTT	TCTAAACAAG	CATCIGAAGT	980	
	TGCCGTTTC	CTCTAAAGC	AGGGATCCC	AGGGCCCTGG	CTGTTGGAGT	GCACCGGCT	GGGGCTGTT	1050
	AGGAACCCG	CGCACAGCGG	GAGGGTAGGT	GGGGTGTGGG	GAGCCAGGT	TCCCCCTG	GCCCCGCC	1120
20	TCTCAGATCA	CGAGTGGCAT	GGGGTGTCTA	GAGGCCAC	CACCTCTAC	AGAAGCTGTC	GTGAGAGGGG	1190
	TCTAGATTCT	GTGCTCTTA	TGGGAATCTA	ATGCCCTATG	ATCTGAGTG	GAACCGTTG	CTCCAAACAA	1260
	CATCCCTCTC	CCCACTGCTG	TCTGTGGAA	AAATCGTCTT	CCACGAAAC	AGTCCTGTT	ACCAAAATGG	1330
	TTGGGACCC	TGTGCTAACG	ACCTGCTCA	GCAGCCTCT	GTCACTGTTG	ATATATTGCG	TTTCTGTGTT	1400
	TGAGTCAGA	ATAATTACGG	ATTTCTGTA	TGCTTCTCCG	CGACCTCAGA	CCCATGGGT	ATTGTGGCC	1470
25	GTGTTGCTG	CTCTGGGT	GGGAAGGGT	CAGGCCCAT	GTACCTTCT	TTAATGCT	TCCAGGTTG	1540
	TTCTCAGGGT	TGAATCGTAC	TCGATGTGTT	TTAGGCCAC	GCCCCCTG	CCAGCTTCC	GGGGCTGGG	1610
	AAACATCTGA	AGCACAGAT	CACCGTGGC	TGCTTTGAT	GCCTCACAAG	CTCGAGGCC	CTCTGTCTC	1680
	TGTTAGTGTG	TGTCACGTG	CTGTCACAT	CTCTGCTTG	GGACGCCAGG	GCTTAGCAGG	TCCCCTAGTA	1750
	AATGACAAGC	GTCTGGGGG	AGTCTCAGA	ATAGGAGGTG	GGGGTGCCTG	TCTCTCTCC	GGCTCTTCA	1820
30	ACTCTCTCC	TGCTGTGCT	TGGCTGAC	TCAGCATCC	CAACCTCTC	CAGCACTGG	CTGGAGAGGG	1890
	CCGGGAGCTG	GAGTCCACT	TGTTGACCGT	GACTGTGGAT	GGCAGTCGGT	CACGGGGGTC	TGATGTGTTG	1960
	TGACTGTGGA	TGGCGTTGG	TCACAGGGT	CTGATGTG	GTGACTGTG	ATGGCGGTC	TGGGGCTGTA	2030
	TGTGGTACT	GTGGATGGGG	GTCGTTGGGT	CTGATGTG	GTGACTGTG	ATGGCGGTC	TGGGGCTGTA	2100
	TGTGGTACT	GTGGATGGGG	GTCGTTGGGT	CTGATGTG	GACTGTGGAT	GGCGCTCTG	GGGTCTGATG	2170
35	TGGTGA	GATGGCAGT	CGTGGGTCT	GATGTTGTT	GACTGTGGT	GGCGCTCTG	GGGTCTGATG	2240
	TGGTGA	GGATGGCAGT	CGTGGGTCT	GATGTTGTT	GACTGTGGT	GGCGCTCTG	GGGTCTGATG	2310
	TGTTGGTACT	GTGGATGGGG	GTCGTTGGGT	CTGATGTG	GTGACTGTG	ATGGCGGTC	TGGGGCTGTA	2380
	TGTTGGTGA	CTGTTGATGG	CGGTGCTGGG	GTCTGTG	GTGACTGTG	ATGGCGGTC	TGGGGCTGTA	2450
	TGTTGGTGA	CTGTTGATGG	TGATCGTCA	CAGGGTCTG	ATGTTGTTG	ACTGTGGATG	CGGGCTGCTG	2520
40	GGTCTGATGT	TGTTGACTGT	TGGATGTTGA	TGGTACACAG	GGGTCTGATG	TGTGTTGACT	GTGGATGGCG	2590
	GTGCTGGGGT	CTGATGTG	GTGACTGTG	ATGGCGGTT	GTGACTGTG	GGCGCTCTG	GGGTGACTGT	2660
	GATGGCGATC	GGTCACAGGG	GTCTGTG	TGGTACGTTG	GGATGGCGGT	CGTGGGCT	GATGTTGTTG	2730
	GACTGTGGAT	GGGGCTGAGT	GGGTCTGAGT	TGTTGACTGT	GTGACTGTG	GGCGCTCTG	GGGTCTGATG	2800
45	GACTGTGGAT	GGCGGCTG	GGGTCTGATG	TGGTACTGT	GGATGGCGGT	CGTGGGCT	GATGTTGTTG	2870
	GACTGTGGAT	GGCGGTTGGT	CCCAGGGGTC	TGATGTG	GTGACTGTG	TGGCGGCT	GGGGCTGAT	2940
	GTGCTGACTG	TGGATGGGGT	TGGGGGGGTC	TGATGTG	GTGACTGTG	TGGCGGCT	GGGGCTGAT	3010
	GTGTTGGTAC	TGGGGATGCC	GTGCTGGGG	TGATGTG	GGTGACTGTG	GATGGCGGTC	TGGGGCTG	3080
50	ATGTTGGTGT	ACTGTGGATG	GGGGCTGCTG	GGTCTGATG	GGTGACTGTG	GATGGCGGTC	TGGGGCTG	3150
	ATGTTGGTGT	ACTGTGGATG	GGGGCTGCTG	GGTCTGATG	GGTGACTGTG	GATGGCGGTC	TGGGGCTG	3220
	GGGCTCTGATG	TGTTGGTACT	GGGGATGCC	GTGCTGGGGT	CTGATGTG	GACTGTGGAT	GGCGGCTG	3290
	GGGCTCTGATG	TGTTGGTACT	GGGGATGCC	GTGCTGGGGT	CTGATGTG	GTGACTGTG	ATGGCGATG	3360
	GTCAACAGGG	TCTGATGTG	GGTCTGACTG	GTGCTGGGGT	GTGCTGGGGT	ATGTTGTTG	ACTGTGGATG	3430
55	GGGGCTGCTG	GGTCTGATG	TGTTGACTGT	GGGGATGCC	GTGCTGGGGT	ATGTTGTTG	ACTGTGGATG	3500
	GGGGCTGCTG	GGGCTCTGAT	TGTTGACTGT	GGGGATGCC	GTGCTGGGGT	ATGTTGTTG	TCAGCTGGA	3570
	GCAGGGTGGAG	TCCCAGGTG	TGTTGACTGT	GGGGATGCC	GTGCTGGGGT	GGGTCTGATG	TGTTGAGCT	3640
60	GAAGCTTCCC	AGGGCTCTC	TGGGCTCAT	CCGGCCATCG	GGGCTGGGG	CAGGTCACA	CTGCTGATC	3710
	GAAGAAACAA	AGTGGCCAGC	TCTGGCCGGG	GCAGGCCACA	TTTGTGGCT	ATGCCCTCTC	CTCTGCGGC	3780
	AGGTCTCTAC	CTTGACAGAC	CTCCAGCCG	ACATGGGACA	TGTCGTTG	CACCTGCA	AGACCAAGCCC	3850
	GTCTGAGGAT	GGCGCTGCA	TGAGGAGGT	CTGGGCACT	CCCTGCA	GGGGTCA	ACCTCCAGCA	3920
65	GTGGGCTCTC	CCCTGGGCAA	TCACTGGGT	CATGACCGGA	CAGACTGTG	GGGTCTGATG	TGTTGAGCT	3990
	GAATGACCTG	TGATGGGGAT	GTGATGAGGT	GTGCTGGCT	GGCAAACTG	TCGCTGGGCA	TGCGAGGCT	4060
	CGACAGCTG	TGCTTCAAGG	CACCTGCTCA	CTGGTCACTG	CGCGGCTCT	CTCCAGTTT	GCAGTGGCTT	4130
	TGTTCATGAT	TTGCTAAATG	TCTTCTCTG	CAGTTTGT	CTTGAGGCCA	AGGGAAAGGT	GTCCCCCTCC	4200
	TTTAGGATG	CAGGGCAGATG	TTGAGGCC	TCTGAGGCC	CTGGGCCCC	AGTGTGGGGT	CTGAGGCCAA	4270
70	AGGAAACCGT	TCCCCCTCT	TAGGAGGACG	GGGCCGTTT	GAGGACCCG	CCGGCTGAGG	GGCCTCTCAG	4340
	TGCTGGGTCT	GTCCACGTG	CCCTGTG	CTTTCAGAT	GTGCTGTC	CCGGCTGCCCC	TGTTGCTCTT	4410
	TGAGATG	TGTTAGCAGT	TGCTCGGCT	TAGGGGACAG	TCTGTC	GGCGATGAGGC	TCAGAGACCT	4480
	CTGGGCAAT	TTCTTGGGT	CCCAGGGTGG	GGGTGGAGGT	GGCGCTGGGT	GCTGGGACCC	AGACCCCTG	4550
	CCCCGCACT	GGGCAGCAAC	TCTGGGATCA	CATATGCCAT	CCGGGCCAGG	GTGGGCTG	TGGGTGAG	4620
	CCCAGCTGGA	CCCACAGGTG	GGCCAGAGGA	GACGTTCTGT	GTCAACACT	CTGCTTAAGC	CCATGTGTTG	4690
	CTCAGAGAC	TCGGGCTGG	CAGCCCCACGA	TGGCCCTGCA	TTCCAGGCCA	GCCCCGACT	TCACTCACAAA	4760
	CACTGACCC	AAAAGGGGAGC	GAGGGTCTG	GCCACGTG	CCTGCTGTC	TCAGCACCA	CCGGCTCACT	4830
	CCCAGTGTG	TCCCCTGCTG	TTTCGAGAG	CTCTCCCTG	AATGAGGCCA	GCAGTGGCCT	CTTCGACGTC	4900
	TTCTTACGCT	TCATGTGCA	CCACGGCG	CGCATCGAGG	GCAAGTGT	CAGGTGGCC	GGTGCAT	4970
	CCCTGCGGGT	GGCTGGCG	GCTGGCAGGG	CTTCTGCTCA	CCTCTCTCCC	GGCCCTCTCC	CACTGNCC	5040

	CTGCCCCGGG	CCACCAAGAGT	CTCCTTTCT	GGCCCCGGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGCTG	GCTTGCGGA	GCCGGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGAAAATGGC	5180
5	AAGCGGGGGT	TGGAGTTGCT	CTGCGCTGG	GGACGAGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCGC	5250
	GAGCGGGTTG	GCCTGCAAGCT	TGTCACTCC	RAGTTACTAC	TGACGGCTGG	CACCCGGCTC	TCACACGCTT	5320
	GTATCTCT	CTCCCGATAC	AAAAGGATT	TATCCGATTC	TCATCTCTGT	CCCTGCGTG	TGACCCCCGC	5390
	GAGGGCCGGG	GCTCTTCTCT	CTGTGACTAG	ATTCCTCCATC	TGAAAGTGC	GGGGTTGACC	GTGTAGTTG	5460
	CTCCTCTCGG	GGGGCCATGG	GTGGCATG	CTGGGAGAGC	TGCCGTACAC	CACCCACTGG	5530	
	GTGAGCCACA	CTCACGGTGG	TAGAGCCACA	GTGCGCTGGT	CCACATCAGC	TCTCTGGAT	TTAAAGTAAA	5600
10	ACACACACCC	TCCCCGGCAGG	CATCTGCTG	CGACCCGTG	TGTGCGCTGG	GAGAGTGGTA	GCACGGAGGA	5670
	AATTCTGTGCA	CACTCAAGGT	CATCAGCAAG	GTCATCCGCA	GTCAGGTGG	ACGCTGGAGG	CTCTCTCTGG	5740
	GATCGCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTCTGG	AGCTTCTTATT	TTAAATATA	ACTATTAAATT	5810
	ATTGCATTAT	AAGTAAATCAC	TAATGTTAC	AGCAATTATA	ATATTATTA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGG	AAAACACAAA	TTGCGCATGG	CAGCGAGACT	RATTTTGGCC	GGAGGACAGC	5950
15	TGTGACATG	TGTGTAAGCG	CCCCCGAGG	CGCACAGATT	CGCTGACAAA	TGACACCTCCC	CAGAGAACGC	6020
	ACACACGGCC	TCCCTCGTGG	TCTGTAATT	TATTAAGATG	GATCAAGTC	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTGG	GAATGTGAGG	TGATGACTGC	GTCCTCATGC	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGCTCCCTAG	GACACGGACA	GGCCCGGAGC	TCTAGTCCC	ATCGTGTG	AGTTTGGCT	CTGAATAAAA	6230
	ACGCTCTCAA	ACCTCTGGC	CCCCAAACT	AGAACACAG	AGAGTTCCC	ATCCCAGTGT	CTCACAGGG	6300
20	CGTATCTGCT	TGCGTGTACT	CGCTGGCTG	GGCCGACTCC	TAGAGTTGGT	GGCGTGTGCTT	CTGTGCAA	6370
	AGTGCAGTCC	TCTGGCCCAT	CACTGTGATA	TCTGCGACAG	CAAGAACAGA	TCTTTCTCT	TTCTTCTTT	6440
	TTTTTTTTT	GAGACGGAA	GTCACTGTTG	TCTGCGCTGG	CTTGAAGTCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GGCTCCGGG	TTCAGCATT	TCTCCGCT	CAGCGCTCCG	AGCAGCTGAG	ATTACAGGCA	6580
25	CCCCACCCCT	GGCGCTGGT	RATTTTGT	TTTTTAGTAG	AGAGGGTTT	TTGCATGTT	GGCCAGGCTG	6650
	GTCTGCAACT	CCTGACCTCA	GGTGTACCC	CCACCTCGGC	CTCCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCACGCC	CAGCGGGAA	GGCTCTTTT	AAGGTGACCA	CCTATAGCGC	TTCCCGAAAAA	TAACAGGTCT	6790
	TGTTTTGCA	GTAGGCTGCA	AGCGCTCTT	ACGACAGGA	GTGCGCTCT	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCCGGT	GGCAGCCATG	CTCTCTGTG	GCACCTT	GTTCCACGGG	GCTATTCTGC	TCTCACTGTT	6930
30	TGCTGAA	GGCACCCTTG	GCATCTGCA	TGAGAGGTT	TCTGCTCTC	GTGGTGTATG	CTGAAACTAG	7000
	GGCGCAAGGTT	GTATGCTG	GGCGCAGGG	GTCATCTGTA	GGGTGATGAG	TCTTCACCG	TGGAACAATT	7070
	CCTTGAAAAA	AAAAAAAGGA	GTCCGGTTAA	GCATTCA	CGGTCATG	GTCTGTTCT	GTGAATAAAC	7140
	TCTAAGGAT	AGAAACCTT	ATAGAAGAA	AACTTGTATG	ATTCAAGGAGA	AGGAATGTTG	CACACCTGTG	7210
	GCTGGATCTG	TTCAGGCCG	CCCACTGCA	GGTGAAGT	GGGAGCAGG	ATGTTTGT	CAAGGGTCTC	7280
	ATCTGTTATG	TTCTGAGGT	TTTGCCTGG	TGAAATGTT	ACGCGTCCTT	TGTGTGATG	AGGTTCTG	7350
35	TCTGTGTTG	AGTGTACGCA	TGTCAGCAC	ATGCGCCAC	CGTCTCTCAC	CTGTGCTTTC	7420	
	CCGCCCCAGG	TCCCTACGTC	GGCGCAGGG	GATCCCCAG	GGGTCATCC	TCTCCACG	GCTCTGTCAG	7490
	CTGCTGCTAGC	GGCAGATGGA	GAACAGCTG	TTTGCCTGG	TTCCGGGGA	CGGGTGGAGG	CTCCTCTTC	7560
	CCAGGGGGC	TGGGGTGGG	GTTGATTG	TTTGTGATCA	TTCAAGTGT	ATATCCCTG	TGCTCTGGAG	7630
40	ACCATGACTG	CTCTGTCTG	AGGAACAGA	CGAAGTGTGCA	GGCCCTCTT	GGTATGAAGC	CCACACGGAG	7700
	GGGTTGCCA	GGCTGAGGAC	TGCCGGCTC	ACCGAGGCTC	TGTCACGGG	CCATGTCCAG	AGGCCCTCAGG	7770
	GCTCAGCAGG	GGGGAGGGCC	GCTCGCTGC	ATGATGAGCA	TGTAATTCA	ACACCGAGGA	AGCACACCA	7840
	CTTCTGTAC	GTCAACCCAGG	TTCCGGTAGG	GTCTCTGGG	AGATGGGCT	GGTCAGGCT	GAGGCCCCAC	7910
	ATCTCCACG	AGGCCCTCTG	CAAGTGGCC	GGAAGTGGGG	CCTCTTCA	CCATTGCCCCA	TCCCATTGTC	7980
45	ATGGGGCTA	CACCCAAGGA	CGCACACCC	TAATATCGT	GCAACCTAA	TGTGGTTCAA	CTCAGCTGGC	8050
	TTTATTGAC	AGCAGTTACT	TTTTTTTTT	TAATCTTAA	ATGCTTCA	TGATGTGCA	CGACGTGCA	8120
	GTTAGTACACA	TATGTATACA	TGTGCGATG	TGTTGTCGCT	CACCCATTAA	CTCATTCA	ACATTAGGTA	8190
	TATCTCTAA	TGCTATCCCT	CCCCCTCCC	CCCACCATCC	GACAGGCC	GGTGTGATG	GTTCCTCCAC	8260
	CTGTGTC	GTGTCTCAT	TGTTCA	CCACCTGTG	GTGAGAACAT	GTGGTGTG	TTTTCTTC	8330
50	CTTGAATAG	TTTGCTCAGA	GTGATGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	TGAACCTCATC	8400
	TTTTTTATG	ACTGCA	AGTCTGCTG	GTATATGTC	CACTATTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTGG	GTGCGTTGCA	AGTCTGCTG	ACTGTA	GTGCCGCAAT	AAACATACCT	GTGCA	8540
	CTTTATAGCA	GCATGATT	TAATCTTAC	GGTATATAC	CAGTAATGGG	ATGGCTGGG	CAAATGGTAT	8610
55	TTCCTAGTCT	AGATCCTTGA	GGAAATCACCA	CACTGCTTC	CAACATGGT	GAACATGTT	ACRACCTCCAC	8680
	CAACAGTGA	AAAGTGTCT	GTTGCTGAG	AGGATGTTG	CAGCAGTT	TTTTTATGA	AAATAGTATC	8750
	ACTGAACAAAG	CAGACAGGTA	GTGAAGGATG	CGTCAGGAGA	CCTGCAAGGC	ACACAGCCAT	TTCCTCTGAA	8820
	GACTCCGGT	TTTCTCTG	CATCTTGA	AACTCTAGC	CCAACTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCAGATT	GGAAATACTT	TATGTAACAG	AAACAAAAAT	TCTCTTAC	8960
	CACAACTTC	TCTGGGATT	GGAGGAAAGT	GTCTGGCTG	CACAGTGGG	ATGGCTGGG	CAAATGGTAT	9030
60	GGATACCTCT	GGCCCATGGT	CATGGGGC	TGGCTCTGG	CCTGAGGGCT	ACACAGTGC	CCATGCCAG	9100
	CTTCTGTG	ATAGGATCTG	GGTCTCGGAT	CGTCTGAGG	ACACAGCTG	CCATGCTG	AAAGGCCACC	9170
	ACCTGGCTCA	GAGGGGGGCGA	GTTCTCCAGC	CCACAGGTTTC	TTACCTGCT	CAGTTATT	TCCCTAAAGAG	9240
	TCTGAGAAGT	GGGGGGCGGC	CTGATGGCT	CTGTTCTCT	TCAGCTGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAAACACTG	AGTACTTATA	ATGAATGAGG	AATTGCTG	GCAGTTA	GTAGAGAGCT	CGTCTGTTG	9380
65	AAAGAAATTT	AAGTTTCTA	TTAACCGCT	TTGGAGAATG	TTACTTATT	TATGCTGTG	TTAAATGTTT	9450
	GACATTCA	CCCTCGTGA	AGATACATC	GTAAAAGTG	TAAGTAAAC	CTGCTG	ATTTTCTCTT	9520
	ATTTAGGCT	GCTCTCTG	TTGGTGGATG	ATTTCTGTT	GGTACACCT	CACCTCACC	ACGCCAAAC	9590
	CTTCTCTG	GTAGGCGCTG	GGCGTGTGTC	TGTTGGGAC	TCCACGCT	GGGGCTT	CAGTTGAGCC	9660
	CCCCGTG	TGCCCCCTGG	ACCCGAGCGT	TGCTCTG	AAGTCTCTC	TCTCTCCGG	TGCTGGATCC	9730
	GCAAGAGCAG	AGGCGCTTGG	CGCTGCA	AGGCGCTGGG	GGCAGGGGC	ACCTTGGGA	GGGAGTGGGT	9800
70	ACCGTGCAGG	CCCTCGTCT	GCAGAGACGC	ACCCAGTTA	CACACGTTG	GAGTCAGGCC	GGTGACCTGG	9870
	CTCTGCTG	TCTTGGAAA	GTCAACAGT	GGGGCTCTG	GGGGCCAGT	GAGACCCCA	GGAGCTGTC	9940
	ACAGGGCTG	CAGGGCCGAG	GGGGCAGCCT	CCTCCCCAGG	GTGACCTGA	GCCTGCGGAG	AGCAGGAGCT	10010
	GCTGAGTGTG	CTGGCCCAACA	GGGTTGGCTG	CGGTCAGGTT	CTCTGGTGGG	GGGGCTT	ATCGGTG	10080
75	GAATTGAT	TTGCTGAGT	CTGCTGCTT	GAACCCAGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
	TTTGTGAA	AAACTAAAAA	CAGGCCACAG	GGACCTGGCC	TCAGCACAGG	GGATTGTC	ATGTGGTCCC	10220
	CCTCAAGGGC	GCCCCACAGA	GCCGGTGGGC	TTGTTTAAA	GTGGATTTG	ACGAGGGAGC	AGAAACCTTG	10290
	AAAGCTG	AGGGAAACCT	CAGAAAATG	GGCCGCCAGG	GGTGGTTCA	GTGCTTTC	TGGGCTG	10360
	TTGTGAAAC	CCATTGGAC	CGGCCCTCA	AGTCCACCT	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGG	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCGGAGC	ACAGCAGGCT	GTGACATT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CTTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGCTG	CTTCTGTGTC	GCAAGTTCT	GAGGGTGTG	GCCAGGGAGG	TGGCTCAGAG	10640
5	TGTATGTGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTG	ATGACTCGGC	AGCCATGTA	CAGGAAGGGG	10710
	TGGCACRGG	GAGCTGGAA	TGCAACAGGG	GAGCTGGCA	GCTGCCAGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGGCCC	GGGCCACAGG	AGAGGGCCGA	GGAAAGGGAG	GGGATGCCA	GGCCAGAGCCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCCA	GGAAACCTCC	10920
	TTGACGTAA	GCTGACGACT	GGTGTGCCCC	AGCTCACAGC	CCAGGCCAGT	CCCGCCTG	AGCAGGAACAT	10990
	CAGAACCTC	CCCTTGTCT	AAAGCACAGC	AGATGCCCTA	AGGGCATCTA	GGAGAAAACA	GGCAAAGTCG	11060
10	TTGAGAACG	TCTTAAAGA	AGGTGGATG	TTGGCAATT	CTTGTCAGA	TTTAGTCG	CCCCGGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCACT	11200
	GGGGCTCCAC	CTCCCATCTG	AGTCTGGCT	GTCCCAGGGT	CAGGCCAGGT	TTCTGCTG	TCACCTACCT	11270
	GTCTCGCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGA	GCAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CTGCCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTCTCT	GGGTTTTC	AAAGCATT	ACAAGGGTGT	11410
15	CAGGTTACCT	CCTGGGTGAC	GGCCTGGGCT	ACATTGCCC	CTGCCTTAG	GACCCCTGGTC	11480	
	CGAGGTGTC	CTGAGTATGG	CTGCGTGTG	AACTTGCGGA	AGACAGTGGT	GAACCTCCCT	GTAGAAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCGGGGCCA	CGGCCATTATC	CCCTGGTGG	GCCTGCTGCT	11620
	GGATACCTGGG	ACCCCTGGG	TGCAAGAGCG	CTACTCCTAG	TGAGGCGACC	TGCGGGAG	TGGAGCCCTGT	11690
20	GCCCCGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTGGCTCC	ACCTCTGCTT	CGCTGTGGG	CAGGGGACTG	11760
	CCAATCCAA	AGGGTCAGAG	GCCACAGGGT	GCCCCCTCGT	CCATCTGGGG	CTGAGCAGAA	ATGCATCTT	11830
	CTGTGGAGT	GAGGGTGCTC	ACAAACGGGAC	CAGTTTCTG	GTGATTTTG	GTAAAAGGAA	ATGGTGACCC	11900
	AGACCTGGGT	GCACTGAGGT	GTCTTCAAGA	AGCAGTCTGG	ATCCGAACCC	AGACGCCCG	GGCCCTGCTG	11970
	GGCGTGAATC	TCTCAAACCC	GAACACAGGG	GCCCTGTGG	GCATGAGTCC	CTCTGAACCC	GAGACCTTGG	12040
25	GTAGGCCCCA	CACTCCAAAG	CTCATCACA	GTCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCC	12110
	TCAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTC	CAGAGCCCGA	12180
	GAGCTCAAGG	CCCCCTCTCA	GGCTCAGACA	CAAATGATTA	GAAGATGCGA	ACAGATGCGA	AAATCTGTC	12250
	TGTTTCTTT	ATGAATAAA	AGTATCAACA	TTCCAGGGAC	GGCAAGGGT	CTCACACCTA	TAATCCCAAGC	12320
30	ACTTGGGAG	GCGGAGGTG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAAC	AACATAGTGA	12390
	AATTCCCTT	CTACTTAAA	AAATCACAAA	TTAGCCTGCT	CTGCTGGCAC	ACGGCTGTAG	TCCCCGCTAT	12530
	GCGGGAGGCT	GAGGAGGAG	AAATCATTGA	ACCCAGGGAG	CAGAGGTTGC	AGTGAAGCGGA	GATCACACCA	12600
	CTGCACTCCA	GCCTGGGCA	CAGAGTGA	CTTCATCTTA	AAAAAAA	AAAAAGTATC	AGCATTCCAA	12670
	AACCATAGT	GACAGGTGTT	TTTTTATCT	GTCTGCTGAT	AAATTTTAC	GTGCTGTG	TAGAGGCCCCG	12740
35	AACTGGGGT	GCCCTTCTCT	GAAGAGCACA	CCTTCATGGG	AGAGAAAATA	AGTGGTGAAT	GGTTGTTAAA	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGCGTC	TGAGTTAAC	GTCCAGATCT	GAACCTTGGC	TCTTCCAGA	12880
	ATGCTCTCTG	GGGGTTGCT	CATGGGGGAG	CAGCAGGTG	GGACACCCCTC	GTGATGGGGC	AGCAGCAGGT	12950
	GCAGACGCC	CTATGATGGG	GGAGGTGCG	GTGCAAGAC	CCTCTGTG	GGTGGCCAGG	ATGCTCCCTG	13020
	TGCACTCCC	TCCCCACAA	GATGCCGTC	TCCCTGTGCTC	CCCACAGTCC	CTGCTTCCC	CTCACAGCT	13090
40	TACCTGGCT	TGGCTTCCAC	TGGCTTTGTC	TGTCATGATT	CCACATTCTT	TGGGCTTCC	GCACCTCTTC	13160
	GCCTCTCCA	GGCACCTCTC	CAGTGTGCG	CATACAGTC	AGCTGTGAC	GTCCACTGC	TTTTTTGCT	13230
	CCCCATGAAA	TGTATTTTT	AGGACAGGCA	CCCTGTGTC	CAGCCTCTGG	CACAGCATCA	GTGAATGTTA	13300
	TTGAAGGACA	AAGGACAGAC	AAACAAATCA	GGAAAATGAG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGCTAGTGC	GGATGGGGT	GCATCAGTC	ATCAGATGTTG	GGTCAATG	CAGAAATTTC	TGTGCTCCCA	13440
45	AAGCCCACTT	GGTCAGAGTG	TGTGCTTGC	GAGGTGGCTC	TTAAAGCTCA	GCAGTGGAGG	CAGTGGTTCG	13510
	CCATACTCAG	GGTGAACCTC	CATCCCTCTG	GTCTGAAGT	TACAGCAGC	GCTGAAGGGG	CTACTGGGG	13580
	AAGAAAACAG	GCAAAATGAT	TAAGAAAAGT	AAAAAGGAA	AGTGGTAAG	ATGGGAATT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCAAACCC	AGCTCAGATG	GTAGAAATG	GTCAAGACTG	ATGGACAGAA	CAATAGAAC	13720
	AAACGGAAAG	CCATCTCTC	AAAGAACCTG	GTTAATGTTG	TATGTTGGC	AGCTGATGGA	AAAGAGATGT	13790
50	TGTGTGAAAT	TTTTTTCT	GAAGAAAAC	ACTGGAAGCA	TTAAATGTTG	GTCTTACAG	CATATACCA	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACA	CCAGCAACAG	AAATAAAACA	AAAGACTCAA	13930
	AGGGAAAGGA	GGTGAACGCT	CCCTGGTTG	GTGTTGGGG	AGGACACAC	GGGAGGGGGA	TGAAACCGT	14000
	GGGCAACGG	GCATTGCTTT	CACTGCAGAG	AAACTCAGCT	TGCTCTGAGC	ACAGTAAAAA	TGGCCATTCC	14070
55	CTGGAGGCTT	TGTGCACTG	ATTTATTTA	GGGCCCTCTG	GAGGTCTGCT	ACATTCTAC	TCTCACTTTG	14140
	TTCTCCAACT	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCAA	GGGGAGCAGC	CGGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGCCA	TGCTGATTG	CAGCCTGCGC	TCTGCTCGG	GGGGCTTGGC	TCTGGCGAG	GACCCACAC	14280
	AAAGTCAGACC	CATAGGCTCA	GGGTGAGCG	GAGGCCAAGG	TGCTGTTGGG	GATGGCTGTG	AAAGAAGRAA	14350
	TGGACGCTG	ATGCACTCT	GGGAAGGTC	TACCCAGTC	TTCAAAAGGAA	TGCAAGAGAA	ACTGACAGCG	14420
60	AGACCCCATCC	CTCAAAAGAA	CGCACGTGAA	ACTGATGGG	AGACCTGTC	CCATCCCTCA	TGCTGGCTC	14490
	TTTCTCTGGG	TTGCCAAAGAG	CCAGCAGTCAG	GTTGAGGCA	GCTGAAAGA	CTTTCTGGA	AAGCAGCTTG	14560
	TTTGATGGA	AGTCCTCACA	ATGCTCTG	TCTTCTCTG	AATTCTCAG	CTGAAGTGC	CAGCACATT	14630
	CACGGGCTT	TTTACCAT	TCCAGTGTTC	CAGGGAGGGG	GACTGCCCCA	ACAAAGTCAC	GAACCTGCCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTAGTCATC	ACAAAAC	CTCTGCCATT	AAACATT	CAAAGAATT	14770
	TTGAGGAATG	TTTAAATGCA	CAAACGTTT	TTTCAATG	AGCAGTGTG	AAAGCTGAT	TGAAAAGAAC	14840
65	ACACCCAGG	AGCCTGGCT	GAATGTGATG	TGTGTTCATC	TTTGAGCATG	GACATACATC	GGCAGTGA	14910
	GGTGGTGAGG	CCCTGGAGGA	CATGGGTGG	ATGCTCTCAT	CCTGCCCTC	TGGAGGACACC	ATGTTGCCCCA	14980
	CGTGCAC	ATGGCCACCT	GTGGAGGCT	TGTCACCTG	CCTCTTCC	ACCTGAGATTC	AAACACAGTG	15050
	AGATTCCTCA	CGCCCAACTC	AGTGTCTTC	ACACAAAC	CTGACTCACA	CTCTGTGTC	CTCAGGGAC	15120
	GCCCCGGACC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTGTG	GCTGAGTTG	GTGCAAGATCT	CATCAGGGCA	15190
70	GATGATGACT	GCACAAACAC	GGCCGTGCGA	GGTTGGATA	CACTCACAT	CACTAGCCAG	GTCTGGTGG	15260
	AGTTGGTCA	TGCAAGACTCT	GGATGGCATG	TAGCATTG	AGTCATG	GTGAGCACCC	AGCCCCCTCG	15330
	GGCTGCACGG	CGAGGCCCCAG	GAAGGACAA	GAAGGGGGAG	GAAGGCAAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTG	GGCAGGAC	CTGCTGCTGA	CATTCCCCC	TGTGTCAG	CTATGCCCCG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTCAAC	CGCGCTTCA	AGGCTGGAG	GAACATG	CGCAAACTCT	15540
	TTGGGCTT	GGGGCTGAAG	TGTCAAGAGC	TGTTCTGGA	TTTGAGGTG	AGCAGGCTGA	TGGTCAGCAC	15610
75	AGAGTTCAGA	GTCAGGAGG	TGTGCGCA	AGTATG	TGTGTCAG	CGCCCGTGG	TGCAAGGCTG	15680
	ATGGTACTG	GCTGCACGTA	AGAGTGCACA	TGTGCGATA	TACACG	CACATACATG	TGTGCA	15750
	TGTACATGAA	GGCATGCGAG	TGTGCGACA	GGTGTGCAAG	GGCACAAGTG	TGTGCA	CGAATGCACA	15820

	CCTGACATGC	ATGTGTGTT	GTGACAGTC	GTGTGGCAT	TCACCGTAGG	TGCATGGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGTAG	GTCCTCAGCA	15960
5	CCAGTGCAC	TCTTACAGG	ATGAGACGGG	GTCCCAGGCC	TTGGTGGGCT	GAGGCTCTGA	AGCTGCAGCC	16030
	CTGAGGGCAT	TGTCCCAC	GGGCATCCG	TCTCTGTG	GCTCTGTG	CCACTCCCC	16100	
	TCTCTGTG	GCATTTACAT	CCACTCCACT	CCCTCTCTG	TGTGGGCATC	CGGGTCACT	CCCCCTCTGT	16170
	GTGGGCATCT	CGGTCCACCT	CCCCCTCTG	TGGGCATTG	CGTCCACTCC	CTCTCCGTG	TCCCTCCGT	16240
	CTTGGCCAG	CTCTGGGGC	AGGCAGATGA	CACAGATG	TGAGCTGGTC	AGGGTGGTC	GCAGCTGCCG	16310
	GGTGAGGGC	AGGCGGATT	TCACTGGAA	GAGGGATAGT	TCTCTGTCAA	AACTGTCCTC	TTCTTGTTC	16380
10	CATCTGAATG	CATGATAAAG	CAAAAAGTAA	AAACTTAAAG	TCCCGAGAG	GTTTCTACCG	TTTCTCACTC	16450
	TTTCTTGGC	ACTCTAGGTG	AAACGCTCTC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCCTGCTGCA	16520
	GCGCTACAGG	TGAGCGGCCA	CCAAAGGGTG	CGACGGCCAG	CTCCAGGGAC	CCTCCGGCT	CTGCTCACCT	16590
	CTGACCCGGG	CTCTCACCTT	GGAACTCTG	GGTTTAGGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTC	16660
	TGCTGCTGT	GCACAGTCT	GTTCGCTG	CTCTGCTG	GGCAACGTTCC	CTGCTCTG	TGTCAGGGG	16730
15	AGGAGCGGCT	GTGGCCCGAG	GTGCTTCCAC	TGTCGCTG	CACTGGCGT	GGGACCGCT	GGAGGACATC	16800
	CCAGGGCAGC	AGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTACG	AGAGGAGGCT	GGGAAGGTTG	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCG	GAGGGATGG	GTCCTCAGCA	AAAGGGCCCG	AGGTGGGTG	16940
	AGGTGAGGGT	CGCTGGCC	ACCCCCGGGA	AGGTGACGCA	GAGCTGTG	TCCCCACACA	GCCCCGGCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCT	GGAACTTACG	CTGCTCTG	TGTCAGGGG	GTGCCCCCTG	17080
20	CAAGAATCGA	CAACTTATC	ACAGAGGGAA	GGGCAAATC	TGTCGCTGAA	CAGGGCAGG	TCTCTGCTG	17150
	AGTCAGGCCA	GGTGGTGGCA	CAAGCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGACCCACA	GGCCCGGCC	17220
	TCCACCTCAA	CAGGCCCTCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATGAGG	17290
	GCTGAGAAGG	AGTGTGAC	TTTGTGTTAC	CCAGGCGCGA	GGCTCGCGCA	ATTACCGTGC	ACACTTGTG	17360
	TGAAATGAGG	TGCTGCTCA	TCTGTTAAC	TCAGCAAGGG	CTCAGGGAG	ATTTTCCAT	TACAAGGTC	17430
25	TACCATGAAA	ATGGTTTTA	ACCCGAGTGC	TTGGGCTTC	ATGCTCTG	AGGGAGGGCA	GAGCCACAGC	17500
	TGATGTTAC	CCCTCTTGC	CAAGCTCCAG	AGGCTGGG	GGAGCTGTC	TCAGTCCCAG	GGTGCGTCCG	17570
	GCTCAGACG	CCCTCTCTC	TGCTTCTCT	CTCTGCTCA	ATCTTCCCT	CCTTGCATC	TCCCCGAC	17640
	GTGCTTGGG	CTCTGTGCAA	GTCCTG	TCTCTTCCGG	AAACCTTGG	GGTGCTG	ATACAGGTC	17710
	CACTGAGGAC	TGGAGGTG	TGACACTG	GTCAGCTG	GGTACCTG	GGGCTCTT	GGGCTCTC	17780
30	GGGCATGAT	AGGTCAGAG	GAGTTTCCC	AGGTGAAAC	TCTCTGAA	TCCCCAGGC	CATGTCACCT	17850
	GCCACCTGCT	CTCTCCAT	TCAGCTCAGT	CTTGTCTCA	TTTCCCAC	AGGGTCTCTA	GCTCCGAGGA	17920
	GCTCCCGTAG	AGGGCCTGG	CTCAGCTG	GGCGGCTGAG	TTTGGG	CATGTCGGG	CCCTTGGGTA	17990
	GTGCTTGTAT	TGGTGTAGCC	TGAGGAGGC	GAGATGCGAT	GGGACACGGG	CCGTTTCAA	ACACAGACTC	18060
	AGGCACGTTG	AGGGCCAGG	AATCCCTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
35	TTTACGGG	GGCAGGCTG	AGTGGCGAG	GTCAGCTG	GTCAGTGG	GCTGCTGATT	18200	
	CAAATCCGCT	GGGGCTCGG	CTTCTCGG	CTGCTG	GGGCTCCAC	ACGGGCTTGG	GGTGGACCC	18270
	CCGACCTCTA	CGAGGTGGCT	ATTCTCTCC	TTGGAAGAGA	GCCCCCTCAC	CATGCTAGGT	TGTTCCCTCC	18340
	TGGGTAGG	GGGGCTGGT	GGTGAACCC	GGGGACCTTA	GGCTTATTTA	TGTTTTAAA	ACATCTG	18410
	GGCTGGCTC	CGTTGTGCT	AAATGGGAA	AAAGACATCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
40	CGGGGGTGTG	GTGTTACTG	TGTTGATCTCA	GGTCATTCCA	GAAGTGGCTG	AGGAAGTCAG	TGAGACCAAGG	18550
	TACATGGGGG	TCCTGAGGAG	GGGGTGGAGAT	GGGGTACACG	GGGGTCTCAG	GGAGTGGGTG	AGGGCAGGTA	18620
	CATGGGGGGC	TCAGGCACTG	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACCAAGTACA	18690
	CGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTGAC	ATGTCGTTG	TCAGTGTAGC	CAGGTCTG	18760
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAG	ATGGGGCTG	ACAGGGCTG	CGCGGTGCT	18830
45	CACACCTGTA	GTCCCAGCAC	TTTGGGAGG	CGAGGGCAGA	GGATCCCTG	AGCCCAGGAG	TTAAGACCA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAAATA	AAAACAAAAA	TTAGCTGAAAC	ATGGTGGTGT	18970
	GCGCTCTAG	TTCCAATACT	TGGGGAGCTG	AAAGTGGGAGG	ATCACTGAG	CCCAAGGAGG	GGAAGCTGCA	19040
	GTGAGCTGAG	ATTGACACC	TGTACTG	CCTGGGTG	AGAGTGGAG	CCCCATCTAA	CAACAACAA	19110
	GAAGACTGAC	AAATGCAAGT	TCTTGGAAAG	AAACATTAG	TAGGAACCTA	ACCTACACAC	AGAAGCCAAG	19180
50	TCGGTGTCTC	GTGTCAGTG	AGATGAGATG	ATGGGCTTC	ACACCATCAC	CCAGACCCA	GGGTTATGC	19250
	ACACAGGGG	CGGGTGGCTC	AGAAGGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	TTGTCG	19320
	GAAGGGCAGG	ATTCACTGTA	AGTACACAG	GGTACACAGA	GAACATGAA	TAAACTGGAA	ACCTTAGAGG	19390
	CCTTCCCGGA	ACAGGGCTA	ATCAGAACCC	AGCATGGGG	GTCGGCATC	AGGATGGAGC	TGCTTCAGCC	19460
	TCCACATGCG	TGTTCATACA	GATGTCGAC	AAACAAAGCAG	TGTCATG	CAACACAGA	CACCGAGCTA	19530
55	CTCGCACACAA	CAAGCACACA	CACAGACATG	CATGCTGCA	TCCGCTG	TGCACTG	CCCATGAGGA	19600
	AACCCATGCA	TGTCGATCTA	TGACGACACA	TGGGGCATG	TGGGGCATG	CCCCACCCA	CGAGCACCGT	19670
	CTGATTAGGA	GGCCTTCTCT	CTGACGCTG	CGGCCATCT	CTCAGGTTTC	ACGGCATG	GCTGCGCTC	19740
	CCATTTCATC	GAAGAACCCCC	ACATTTTCC	TGCGCTCAT	CTCTGACACG	GGCTCCCTCT	19810	
60	GCTACTCCAT	CTGAAARGCC	AGAACGAGCAG	GTATGTG	GTGCGCTGCC	TCAGTGGCAG	CAGTGGCTGC	19880
	CTGCTGGTGT	TAGTGTGTC	GGAGACTGAG	TGAATCTGG	CTTAGGAAAGT	TCTTACCCCT	TTTGGCATCA	19950
	GGAACTGGT	TAACCCAAAC	ACTGTCAGC	TCGTCG	GGCCCTCTG	GGGGTGGAGCA	GAGCACCTGA	20020
	TGGAAGGGAC	AGGAGCTGTC	TGGGGAGCTG	CATCTCTCC	ACCTCTGCT	GCTGTTGGGAA	GGCTGGGG	20090
	GCCTGGTCTC	TCCTGTTG	CCCCATGGT	GATTGGGGG	GGCTGGCTC	TCTGTTG	CCTGTTGG	20160
	GATTGGGTG	TCTCCGGT	ATGGCTACTA	GGGGCTCTG	GGGACACCC	GCAAGGGGCT	TAGGAGGAGG	20230
65	CCAGGGCCAG	GTACCCACAC	CCCCCTCAGG	AGCAGAGGCC	GGCTGATC	ACACAGAGC	CCCGGGCGT	20300
	CCTCTGCTC	CCAGTCACCG	TCCCTG	CTGGACACT	TGTCAGCAT	CAGGGAGGTT	TCTGATCCG	20370
	CTGAAATTC	AGCCCATGTC	AACTGCGGT	CCTGAGCTTA	ACAGGCTCTA	TCTTACCTG	TTTGTGTTG	20440
	GTGGAATTT	CACCTGGAGA	AGCGGAAGAA	AAACATTCTG	TGTCAGTCT	GCTGTTGGT	GGGTGGGAC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCG	CGGGTGTGGG	CAGCTTCCG	GTGTCCTG	GGAGGGGAGC	20580
70	TGGGGCTGGC	CTGTGACTCC	TGGGGCTCTG	TTTCCCCCA	GGGGATGTCG	TGGGGGCCAA	GGGGCCCGCC	20650
	CCCCCTCTG	CTCTGGAGG	CGTGCAGTGG	GGTGCAC	AAGCATTCTC	GCTCAAGCTG	ACTCGACAC	20720
	GTGTCACCTA	CGTGCAC	CTGGGGT	TCAGGACAG	CAAGTGTGGG	TGGAGGCGAG	TGCGGGCCCC	20790
	ACCTGCCCCAG	GGGTCACTCT	TGAACGCCCC	TGTCG	GGGAGGGGAG	ACGAGCTCTA	GTGCTG	20860
	CCCCCCCCGGC	CTGACCTCTG	GGGCTGGAG	CCACCGTGG	AGCCCTATGT	GATTTAAACG	TGGTGTCCCC	20930
75	AGGCCACGGA	GCTGGCAGG	GTCCCCAAT	TCTTGACCC	CTGCTTCCA	TCTCAGGGG	GATGGCTCCC	21000
	CACCGCTGGG	AGGCTCTCTG	CCCCCTGACT	TCAGGCTCT	TCAGGCTCTT	CCCTGGGCTG	TGCCCCGAGC	21070
	TCTCTGGGTC	CTGAGCAAGT	TCTCTCCCG	GGGGCTGGCT	CCAGGCTAC	TGGGCTGCT	GTCTGCTCCC	21140
	CCCCGGTGGAG	GGGTGTCCTG	CCCCTCACG	AGGTTCCCAC	CAGCCAGGGC	CACGAGGTGC	AGGCGCTGCC	21210

	TGCCCCGCCA	CCACACGTC	CTAGGAGGGT	TGGAGGATGC	CACCTCTGGC	CTCTTCTGGA	ACGGAGTCTG	21280
	ATTTTGGCCC	CGCAGCCCAG	ACCGCAGCTGA	GTCGGAAAGCT	CCCGGGGAGC	ACGCTGACTG	CCCTGGAGGC	21350
5	CGCAGGCAAC	CCGGCACTGC	CCTCAGACTT	CAAGGAGCT	CTGGACTGAT	GCCCCACCCG	CCACAGCCAG	21420
	GCCGAGAGCA	GACACCCAGCA	GCCCTGTAC	GCCGGGCTCT	ACGTCCAGG	GAGGGAGGGG	CGGCCCCACAC	21490
	CCAGGCCCCG	ACCGCTGGGA	GTCAGGAGGC	TGAGTGAGTG	TTTGGCCAGG	GCCTGCATGT	CCGGCTGAAG	21560
	GCTGAGTGTC	CCGGCTGAGGC	CTGAGGAGT	GTCCAGGCAA	GGGCTGAGTG	TCCAGCACAC	CTGCCGTCTT	21630
	CACTCCCCCA	CAGGCTGGGG	CTCGGCTCCA	CCCCAGGGCC	AGCTTTCTC	CACCAAGAGC	CCGGCTTCCA	21700
	CTCCCCACAT	AGGAATAGTC	CATCCCCAGA	TTCGGCTATTG	TTCACCCCTC	GCCCCGCCCT	CCTTIGGCC	21770
	CCACCCCCAC	CATCAGGTG	GAGGAGCTGA	GAAGGAGCTCT	GGGAGCTCTG	GGAAATTGGGA	GTGACCAAAAG	21840
10	GTGTCGCTG	TACACAGGG	AGGACCTCTG	ACCTGGATGG	GGGTCTCTGT	GGGTCAAATT	GGGGGGAGGT	21910
	GCTGTGGGAG	TAAAATACTG	AATATATGAG	TTTTTCAGTT	TTGAAAAAAA	TCTCATGTT	GAATCTTAAT	21980
	GTGCACTGCA	TAGACACCAC	TGTATGCAAT	TACAGAACG	TGTAGTGAAG	GGGGTGGGTTG	GTGAGTGCAGG	22050
	GCCCCATGCC	TGGCTGTGCA	TTAACGAGAG	TCTATGAGTG	ATGGGGTTG	TGTCAGTGC	GGGCCATGG	22120
	CCTGGCTGGG	CCTGGGAGGT	TTCTGATGCT	GTGAGGAGG	AGGGGAAGGA	GGGTAGGGGA	TAGACATGG	22190
15	GAGCCCCCAC	CCTGGAAAGAC	ATAACAGTAA	GTCCAGGAC	GAAGGGCAGC	AGGGATGCTG	GGGGCCCCAGC	22260
	TTGGGCGGG	GGGATGATGG	AGGGGCTGGC	CAGGGTGGCA	GGGATGATGG	GGGCCCCAGC	TGGGGTGGCA	22330
	GGGGTGTG	GGGGGGCTGG	TCTGGGTGGC	GGGAAAGATG	GGGAAGGCTG	GCTGGGCCCC	CTCCTCCCT	22400
	GCCTCCACAC	TGCACCGCTG	GATCCGGATG	TGTCACATCC	TCTGGGCTAT	CAGCTTTCAT	22470	
	GGAGGTGGGG	GGCAGGGGCA	TGACACCATC	CTGTATAAAA	TCCAGGATTC	CTCCTCTGA	ACGCCCAAC	22540
20	TCAGGTGAA	AGTCACATTC	CGCCTCTGGC	CATTCTCTTA	AGAGTAGAC	AGGATTCTG	TCTCTGAAGG	22610
	GTGGGTAGG	TGGGGCAGTG	ACACAGGG	CTTCAGGGTG	GGGGTGTGGA	TGCTCTCTCA	22680	
	TCTCTTATC	ATCTCCAGT	CTCATCTCTC	ATCTCTTAT	CATCTCCAG	TCTCATCTG	TCTCTCTTA	22750
	TCTCCCAGTC	TCATCTGTCA	TCCCTTAC	ATCTCCAGT	CTCATCTCTT	ATCTCTTAT	TCTCTAGCT	22820
	CATCCAGACT	TACCTCCCA	GGCCGGTGC	AGGCTCCAG	TGGAGCTGAG	CATACTCTCT	TCTCTAGGCA	22890
25	GAAGGAATCG	GAAGGATTG	AGAGAACAGG	AGGGGGCGCT	CAGAGGGAGG	CAGTCTGGG	GTGAAGAAC	22960
	AGCCCCCTCT	CAGAAGTTGG	CTTGGGCCCC	ACGAAACCGA	GGGGCCTGCG	TAAGTGGCTC	CAGAGCCTTC	23030
	CAGCAGGTCC	CTGGTGGGGC	CTTATGTTAT	GGCCGGGTG	TACTGAGTGC	ACCTTGGACA	GGGCTCTGG	23100
	TTTGAGTGC	GGCCGGACGT	GGCTGGTGTG	GGGGTGGGGG	CTTATGGCCA	CTGGATATGG	CGTCATTAT	23170
30	TGCTGCTGCT	TCAGAGAATG	TCTGAGTGC	CGAGCCTAAT	GTGTATGGT	GGCCCAAGTC	CACAGACTG	23240
	GTCGTAATG	CACTCTGGTG	CTGAGTGCAG	CCGTATAGGA	GCTGTGAGGA	AGGAGGGGGT	CTTGGCAGCC	23310
	GGCCCTGGGG	GGCCCTTGGC	CTGCAAACG	GAAGGGAGG	GGCCGGGGG	CGCTGGGGG	ACGACCTCAA	23380
	GTGAGAGGTT	GGACAGAAC	GGCGGGGAC	TTCCCAGAG	CAGAGGCCG	TGCTCAGGCA	CACCTGGGTT	23450
	TGAATCACAG	ACCAACAGGT	CAGGGCTACG	TTAGCTATC	AAGAGCTCCAG	ATTCTCTTT	23520	
35	CTCGGGTGT	TTTTTGTG	AAATTCTACT	AGGATTACTT	ATATTTTG	CTAAAGTATT	AGACCTTAA	23590
	AAAAGGTATT	TGCTTGTATA	TGGCTTAAC	CACTAAGC	CTACTTTATT	TGTCTGTTT	TATTTATTAT	23660
	TATTTATTAT	ATTAGAGATG	TGTCAGTACTC	TGTCACCCAG	GTGTTAGTG	CAGTGGCACA	GTATGGCTC	23730
	GCTGTAGCG	CAAAACCCCA	GGCTCAAGT	ATCTCCGGC	CTCAGCTCC	CAGAGTGTG	GGATTACAGG	23800
	TGTGAGGCCAC	TGCCCTTGC	TGGCACTTT	AAAACCAC	ATGTAAGGTC	AGGTCCAGTG	GCTCCACAC	23870
40	CTGTCATCCC	AGTAGTTGG	GAAGCCGAGG	CAGAAGGGT	GTCTGAGGGC	AGGAGTTGA	GACCAGCATG	23940
	GGTAACATAG	GGAGACCCCA	TCTCTACAA	AAATGCAAA	AGTTATCGG	CGCTGGGGTC	CAGCATCTG	24010
	AGTCCCAGCT	GCTCGGGAGG	CTGAGTGGGA	GGATCGCTTG	AGCCCGGGAG	GTATGGCTG	CAGTGAGCTG	24080
	TGATTGTACC	ATCGCACTCC	AGCCCTGGCA	ACAGAGTGAAG	ACCCCTGTCT	AAAAAAAAG	AAAAAAAAG	24150
	AAGGAGAAGG	AAAGAGAACG	AAAGGAAGAG	AAAGGAAGG	AAAGGAAGG	AAAGGAAGG	AAAGGAAGG	24220
45	AAGGAGGCCCT	GCTAGGTGCT	AGGTGACTG	TCAAACTC	GAGAAAATG	AAAATAACAA	AGTTTAAG	24290
	GGAAAGAAAAA	ACCCACGTC	TTTGGACTTC	CTTAGGGCTG	AACCTCATCT	CAAGCAGCTT	CCTTCCACAG	24360
	ACAACGCTGT	ATGGAGCCG	TGAGTTCAA	GCAGAAAAGG	AGGAAGAGCA	GGCAAGGGT	GAGGCTGTGG	24430
	GTGACACCAG	CCAGGACCCC	TGAAAGGGAG	TGGTTGTTT	CCTGCTCAG	CCCCACGCTC	CTGCCGTCTC	24500
	TGACACCTGCT	GTAAACCGTC	ATGTTGGTG	CAGGTGCCCA	CCTGGGAAGG	ATGCTGTGCA	GGGGGCTTGC	24570
50	CAAACTTGG	TTGGTTCA	AAAGCCCCAGG	CACCTGTGCG	AGGCACAAAT	ACAGCCCCCT	CCCCAAAGATG	24640
	CCCCACGCTCT	TCTCTTGGAA	CTCTGAATG	TGTCACTCCG	AAGGCAGAGG	CTGGTGAAGG	CTGCAGCTGG	24710
	AATCACCGCT	GCCAGTCAGC	CGATCTTAAG	GTCTACCTTG	ATTATCTGGT	GGGCTGTATA	TGCCCACAG	24780
	GGTCCCTAGA	AGTGAGAGAG	GGAGGGAGGG	GAGAGTCAGA	GAGGGGAGGT	GAGAAGGACC	ACTGGCCACT	24850
	GCTGCTTGTG	AGATGGAGGA	GGGGGTCCCC	AGCCAAGGA	TGGGGCAGC	CGCTCCATGC	TGAAAAGCA	24920
55	AGCAATCTC	CCCCTGCCCTG	AGGGCACACG	GCCCTGCCA	CGCCGAGG	TCAAGGCCAGT	GGGACCTGTT	24990
	TCAGCTTCC	GGCCCTCCAGA	GGCTGTAAGAT	GATGCGTTG	TGTTCAAGCA	CTAAGCTGCA	GTGATTGCTC	25060
	ACAGCAGCAA	ATGGAATAGC	AGTACAGGGA	AATGAATACA	GGGACAGTTC	TCAGAGTGC	TCTCAGCCCA	25130
	CCCCCTGGG							25138

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

65

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon					Intron				3'-Exon						
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk.(%)	70	60	80		100	100	95	70		80	100	100	60		

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

- 5' Intron 1: bp 11493-11596 (SEQ ID NO 4);
Intron 2: bp 12951-21566 (SEQ ID NO 5);
Intron 3: bp 21763-23851 (SEQ ID NO 6);
5 Intron 4: bp 24033-24719 (SEQ ID NO 7);
Intron 5: bp 24900-25393 (SEQ ID NO 8);
5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden
10 Positionen lokalisiert:

- 3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);
Intron 7: bp 3879-4858 (SEQ ID NO 11);
Intron 8: bp 4945-7429 (SEQ ID NO 12);
Intron 9: bp 7544-9527 (SEQ ID NO 13);
15 Intron 10: bp 9600-11470 (SEQ ID NO 14);
Intron 11: bp 11660-15460 (SEQ ID NO 15);
Intron 12: bp 15588-16467 (SEQ ID NO 16);
Intron 13: bp 16530-19715 (SEQ ID NO 17);
Intron 14: 19841-20621 (SEQ ID NO 18);
20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCGGGTCTGGCTCGGCTGGCTGGGTTGAGGGCGCCGGGGAAACAGCGACATGCGAGAGCAGCCAGG
CGACTCAGGGCGCTCCCCCGAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCCAGGCCAGAGCTGAATGCACTAGGGCTCAGAAAAGGGGCAGGCAGAGCC
CTGGTCTCTCTGTCTCCATCGTCAGTGCCACACGTGGCTTCGCTCAGGACGTCGAGTGGACACGGTATCTGCC

10

TCTGCTCTCCCTCTGTCCAGTTGCATAAAACTTACGAGGTTCACCTCACGTTTGATGGACACGCCGTTCCAGGCGC
CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGCGCCAGTGGAGCCGGTTGCCGCAATGGGAGAAGTGTCTGGAAG
CACAGCCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTCTCGCAATTCAAGGGTGGGAATGAGAGGTGGGA

CGAGAACCCCTCTTCTGGGGTGGAGGTAAGGGTTTGAGGTGACAGTGGTCAGTGGTCAGCCAATATGAGGGTTGTTA

AGATTTAATTGTGTGTTGACGCCAGGTGGCTCACGCCGTAATCCAGCAGTGGGAAGCTGAGGCAGGTGGA

TCACCTGAGGTCAAGGAGTTGAGACCAGCTGACCAACATGGTAAACCCATCTGTACTAAAAATACAAAATTAGCTG

15

GGCATGGTGGTGTGCTGCTGAATCCAGCTACTTGGAGGCTGAGGCAGGAGAACTACTTGAACCAGGAGCGAGGC

TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGCGACAAGAGTGAACACTCTGTCTTTAAAAAAAAAGTGT

CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCAGCACAGATCCTGGTCCATTTAGGTAT

GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCACCTGCTGAGGAAGGGACAGTGTGTTGGGTGTTCAAGGG

ATGGTGTCTGGGCCCTGCCGTGCCCCACCCCTGTTTCTGGATTGTGTTGAGGAACCTCCGCTCAGCCCCCTT

20

TGGCTCCCACTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCGATTTCAACCCCTCCCCAACAAACTCCAAAGAC

ATGTAAGACTCCGGCCATGCAAGACAAGGAGGGTGACCTTCTGGGGCTCTTTTTCTTTTATGGTGGC

AAAAGTCATAAAACATGAGATTGGCACTCTAACACCGTTCTGTGTACAGTGCAGAAATTGCTAACTCGCCGGTGT

CAGCAGGTTGCTGAAATGCTGCGTCTTGTGACTGGAAAGTCCCTACCCATCGAACGGCAGCTGCTCACACCTGCTG

GGCTCAGGTGGACCACGCCAGTCAGATAAGCTCATGCAACCCAGTTTGTCTTGTGCTCCAGCTCCCTGTTGAG

GAGAGTTGAGTTCTGTGTCAGGACTCTGCTGTATTGCTCTGACTTCAGATGAGGTACAATCTGCCCTGG

CTTATGCAGGGAGTGGCTGGTCCCCGGGTGCTCCCTGTCACGTGCAAGGTGAGTGGCGTTGCCCTGGAGGTG

GTCACGTGAGGGTGAGTGAGGCGCGGCCGGGTGCTCCCTGTCACGTGCAAGGTGAGTGGCGTTGCCCTGGGTGT

CCCTGTCACGTGAGGGTGAGTGAGGCGCCATCCCCGGGTGCTCCCTGTCACGTGAGGGTGAGTGGCGTTGCC

GTGTCCCCGTGCCCCGTGAGGGTGAGTGAGGCACTGTCCCCGGGTGCTCCCTGTCACGTGCAAGGTGAGTGGCG

CCGGGTGCTCCCTCAGGTGAGGGTGAGTGAGGCGCGGCCAGGGTGTCCCTGTCACGTGAGGGTGAGTGGAG

GTCCCCGGGTGCTCCCTCAGGTGCAAGGTGAGTGAGGCGCTGTCCCTGGGTGCTCCCTGTCACGTGCAAGGTGAGT

GAGGTCTGTCCCCAGGTGCTTGGCTTCAGGTGAGGGTGTCCCTGTCACGTGAGGGTGAGTGGCGTTGCCCTGGGTGAGT

GCGCCGGTTGCCATTGCTGGTAGATGGTGAGGCCAGTGTGCTGGTCCCAAGCTATCTTTCTGATGCTCGCT

25

TCTTGGTCACCTCTCCGTTCCATTGCTACGGGACACGGGACTGCAAGGCTCTGCCCTCCGCGTGCAGGCCACTGCA

CCACAGCTTCAGGTCGGCTTGCTCTGTGGGCTGGCTGCTCACACGTGCCAGCATGCTGCAATACTCC

TCTCCCACTGCTCATGCGAGGCTGGACTCTGGGCTGCCCTGTGCTGCTGCCACGTGTTGCTGGAGACATCCCAGAA

AGGGTTCTGTGCCCCGTGAGGGAAAGCAAGTCACCCAGCCCCCTCACCTGTCTGTTCTCCCAAGCTGCCCTG

TTGGCCCCCTGGGGGGGGCAACGCTGTGACCTTATTCTGGCACCTGCCGCTATTGCTTAGGCTGGCTCGCT

30

CCAGTCGCCCTCACATGGATTGACGTCAGGCCACAGGTGGAGTGTCTGCTGCTGCTGCTGAGACCCACGTG

35

TCTTGGTCACCTCTCCGTTCCATTGCTACGGGACACGGGACTGCAAGGCTCTGCCCTCCGCGTGCAGGCCACTGCA

CCACAGCTTCAGGTCGGCTTGCTCTGTGGGCTGGCTGCTCACACGTGCCAGCATGCTGCAATACTCC

TCTCCCACTGCTCATGCGAGGCTGGACTCTGGGCTGCCCTGTGCTGCTGCCACGTGTTGCTGGAGACATCCCAGAA

AGGGTTCTGTGCCCCGTGAGGGAAAGCAAGTCACCCAGCCCCCTCACCTGTCTGTTCTCCCAAGCTGCCCTG

TTGGCCCCCTGGGGGGGGCAACGCTGTGACCTTATTCTGGCACCTGCCGCTATTGCTTAGGCTGGCTCGCT

40

GAGGGCCGGTCTCGCAGCCTCGTAGACTCCCTCTGGGTCTAGTTTGAAATTCACTGATTTACCTCTGAC
TTCTATCTCCTGATGCTTTCTGGTTTATTCTTCTGTTAGCTTCTAGTTAGTCATGCCCTTC
CCTCTAAGTGCCTACCTGCACCCCTGTGTTGATGTAAAGTAATCTCAACATCAGGCCACTTCAAGTGTCTAAA
ATACCTCAAGTGTAAACTCTTTAAGTATTCTTCTGTGATTTCCTTCTGACGCTGTGTTGACGTGA
5 AATCATTTGATATCAGTGAACCTTAAGTATTCTTCTGTGATTTCCTTCTGAGCAGTGAGTTATTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACTGAGTATTAAAGTTATCATTATTGATTCTAACACTGAGT
TGTGAGTGGTCTGTATAATACCAATTATTGAAGTTTGCAGGCCTGCTTGTGATCTAGTGTGTCATGGTTCCAG
AACTGTCCATTGTAATTTGACATCCTGCAATGTTGAGTCACTATCCAGTTATTAAAGTCCAGTGCA
AAGCTTCTGTCCTCTAGATGCAATTCCAAGAAGGGCCATGTCCTCACCTGGGGATGGTCTGTCATT
10 TCTCTCGTTGGTAGCATTATGTGAGGCATTGAGGTGATGACCTGGTAGAATTATCTCTGATGAGTGAA
TCTTTGGAGACTTCTATGTCCTAGTAATCTAGTAATTCTTTTAAATTGCTCTAGTACTGCCACACTGGGCTCT
TTTGATTAGTATTCTCTGCTGTCGTCTGTTCTGCCTTAATTATATATATATATATTGAGACA
GAGCTTGGTCTGTCGCCAGGGTAGTGCAGTGGTAGTCACAGGTAGTAACTTTACCTCTGGCTGAGCCGT
15 CCTCTCACCTCAGGCCCTGAGTAGCTGGAACCTGCAGACACGCCACCGTACACCTGGTAATTAAATTCTGGAA
GACAGGGTCTGCTGTTGCCAGGCTGTCAAACTCTGGACTCAAGGGATCCATCACCTGGCTCCAAAGTG
CTGAATTACAGGCATGAGCACCAGTCTGGCTTAATTCAACACTTTATATTCTATAGTGTGGTATGTCCTGTA
ACAGCATGTAGGTGAATTCCAATCCAGTGCAGTCGTTAACTGGATAACCTGATTATTCAATTGTC
ACTAGAGACCCGCTGGTCACTCTGATTCTCCACTGCTGCTGTCATGTCCTGGCTTCTGTTCTACCCACCTGG
GGTTGCCATGCGTTCTGCCAGGTGTTGATCTCTGCTGCTGCTGACTGGCATTGCTTTATTCT
20 CTTGCTTAGTGTACCCCCCTGATCTTTATTGCTGTTGTTGCTTTATTGAGACAGTCTACTCTGTCACCCA
GGCTGGAGTGAATGGCACAATCTGGCTACTGCAACCTCTGCCCTCGGTCAGCAGTCTCATCCCTCAACCTCA
TGAGTAGCTGGATTACAGGCCACACCACGCCGCTGCTTAATTGAGTACTGGCTTCTGCTTCTGCTTCTG
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGTACTGCCGCTGGCTCCACAGTGTGGATTACAGGTGAA
GCCACCGTGCCGGCATACCTGATCTTAAATGAAGTCTGAAACATTGCTACCTTGCTGAGCAATAAGACCTT
25 AGTGTATTAGTCTGGCACCCCCCAGCCTGTTGCTGTTCCCTGCTGACTTAGTCTATCTCAGGCATCTGACA
CCCCCACAAGCTAAGCATTATAATATTGTTCTGGTAGTGTCTGAGCTTCTGAGCTTGGCCCTGCCGCTTCT
TTTGTCTCCCGTCTGCTCTGTCAGGCCCGCTGTTGGGCTCCCTCTGCTTGTCTTGTGTTCTGCTTCTG
TTATTGCTGGAAACCCCAGCTTACCTGCTGGCTCCATGGCATTAGCGACGTCGGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCAGGAGGGCGGTACTCTGGCCCGTGAGCTGGAGCACCACGTGGCCAGGTT
30 CTTAGCCAGTGTAGTGAACAGCAACGCTCCGCTGGCTGGTTGAGCTGGCTGGAAAACCCCAGGCATGTCGGGCTGCTG
CCGGTGTGAGTTGAAATCGCGAACCTCGCGGTGGCGCAGCTCTGACGGTGCTGCTGGGGGGAGTGTCTG
CTTCTCCCTCTGCTGGAACCCAGGACAAGGATGAGGCTCGAGGCCAACAGGAGCATGACGTGAGC
ATGTGGATAATTAAATTCTAGGCTGGCGCGTGCAGCCTGTAATCCAGCATTGGAGGCCAAGGGGG
TGGATCACGAGGTCAAGGAGGTGAGACCATCTGGCAACATGATGAAACCCCCTGTAATTAGTACT
35 TGGGCTGGTGGGGTGGCTGTAATCCAGCTACTGGAGGCTGAGGAGGAAATTGCTGAACTGGAGTTGGAA
GTTGCAGTGAACCGACATTGCACTGCACTCCAGCCTGGCAACACAGCAGACTCTGTC
AAAAAAAAAAATTCTAGTAGGCCACATTAAAAAGAAAAAGGTAATTAAATGAAATTAGATTACTGAA
GCCAGCAGTGCACACCTCATCTTAAAGGGTTATTGGTGGAGCATCACTCACAGGACATTGACATTGTC
TTTGTCTGGGGATCCGTGTTAGGCTGGCTGGCATCTGGCTGGACCTGCTGGCTTCCATGGCATGGCT
40 GTTGTACCAAGATGGTGCAGGCTGGGATGAGGTCGCCAGGCCCTGCTGAGCTGGATGTTGCTGAGTGTG
TCTGGGATGAGGTGCCAGGCCCTGCTGAGCTGGATGTTGCTGAGTGTGCTGGGAGGGTCAAGGCTCAG
TCTGGGATGAGGTGCCAGGCCCTGCTGAGCTGGATGTTGCTGAGTGTGCTGGGAGGGTCAAGGCTCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGAT
 TGTGGTGTCTGGATGGTCAGGTCAAGGGTGAGGTCTCCAGGCCCTCGTAAGCTGGAGGTATGGACTCCGGATGATGCA
 GTCAGGCCCTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTCAGGTCTGGATGGTGAGGTGCCAGACGGTGCCAGACCATGC
 5 AGGCCCTGCGGTGAGCTGGGTGCGGTGCTGGATGGTCAGGTCTGGATGGTGAGGTGCCAGACGGTGCCAGACCATGC
 GGTGAGCTGGATATGCCGTCCGGATGGTCAGGTCTGGGTGAGGTGCCAGGCCCTGCTGTGAGTTGGATGTGGGT
 GTCCGGATGCTGCAGGTCCGGTGAGGTCAACAGGCCCTGCTGTGAGCTGGATGGTGAGGTGCCAGACGGTGCCAG
 GGGGTGAAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGAGGTGCCAGACGGTGCCAG
 10 GCCCTCGGTGAGCTGGATGGTGAGTCAGGTGAGGTGCCAGGCCCTGGGTGAGGTGCCAGACGGTGCCAG
 TGCGGTGTCTGGATGGTCAGGTCTGGAGGTGCCAGGCCCTGGTGAGCTGGATGTGGTGAGGTGCCAG
 GGTCCGGGTGAGGTGCCAGACCCCTGCTGTGAGCTGGATGGTGCGGTCTGGATGGTACAGGTCTGGAGTGAGGTGCC
 AGACCCCTGCTGTGAGCTGGATATGCCGTCCGGATGGTGAGGTGCCAGGCCCTGCTGTGAACTGGATGTGGCG
 GGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGAGGTCTCC
 GCAGGTCTGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGAGGTATGGAGTCAGGTCCGGGTGAGGTGCCAG
 CCAGGCCCTGCTGTGAGCTGGATGTGCCGGTCTGGATGGTGAGGTCTGGGTGAGGTGCCAGGCCCTCG
 15 GAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG
 GTGCAGTCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGCTGTGAGGTCTGGATGGTGAGGT
 CACCAAGGCCCTGCGGTGAGCTGGTTGTGCCGGTCTGGATGGTGAGGTCCGGGTGAGGTGCCAGGCCCTCG
 TGGATGTGCCGGTGTCCCCGTGCTGGATGGTGAGGTGCCAGGCCCTGCTGTGAGGTCTGGATGGTGAGGT
 20 GTCCGGATGGTGAGGTCTGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGCTGTGAGGTCTGGATGGTGAGGT
 GGTCCGGATGGTGAGGTCTGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGGTGCTGTGAGGTCTGGATGGTGAGGT
 CCGGGGTGAGGTCAACAGGCCCTGCGGTAGCTGGATGGTGAGGTGCCAGGCCCTGCGGTGAGGT
 CCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGAGGTCCGGGTGAGGTGCCAGGCCCTGCG
 25 TGCTGTATCCGGATGGTGAGGTCTGGGTGAGGTGCCAGGCCCTGCGGTAGCTGGATGGTGCTGTGAGGTCTGGATGGTG
 GGTCCGGGTGAGGTCAACAGGCCCTGCGGTAGCTGGATGGTGCGGTCTGGATGGTGAGGTCTGGATGGTGAGGT
 AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGAGGTCCGGGTGAGGTGCCAGGCCCTGCG
 ATGTGCTGTATCCGGATGGTGAGGTCTGGGTGAGGTGCCAGGCCCTGCGGTAGCTGGATGTGAGGTACGGATGG
 30 TGCAGGTCCGGGTGAGGTGCCAGGCCCTGCGGTAGCTGGATGGTGAGGTGCCAGGCCCTGCGGTAGCTGGATGGTGAGGT
 CGCCAGGCCCTGCGGTGAGCTGGATGTGCGGTCTGGATGGTGCTGCAGGTCCGGGTGAGGTGCCAGGCCCTCG
 TGGATATGCCGTGCTGGGTGAGGTGCCAGGCCCTGGATGGTGAGGTGCCAGGCCCTGGATGGTGAGGT
 GTCCGGATGGTGAGGTCTGGGTGAGGTGCCAGGCCCTGGATGGTGAGGTGCCAGGCCCTCG
 GGTGAGGTCAACAGGCCCTCGGTATGGATGTGCCAG
 35 Intron 3 (SEQ ID NO 6)
 GTACTGTATCCCCACGCCAGGCCCTGCTCTCGAAGTCTGGAACACCCAGGCCGCTCAGCATGCCCTGTCTCCACT
 TGCCCTGTCTCCCTGGCTGTGCAGCTCTGGCTGGAGCCAGGGGCCCTCACAGGCCCTGGTCAAAGTGGATTCTGTG
 CAAGGCTCTGACTGCCCTGGAGCTCACGTTCTTACTTGTAAAATCAGGAGTTGTGCCAAGTGGTCTTAGGGTTGTA
 AAGCAGAAGGGATTTAAATTAGATGGAAACACTACCAACTAGCCCTCTGGCTTCCCTGGATGTGGCTGATTCTCTC
 40 TCTCTTTTTTTCTTTGAGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAGTGGCATAATCTGGCTCACT

GCAACCTCCACCTCCTGGGTTAACCGATTACCGCCTAGCCTCTAAGTAGCTGGATTACAGGCACCTGCCACCAC
 GCCTGGCTAACTTTGTACTTTAGGAGAGACGGGGTTCACCATGTTGCCAGGCTGGTCTCGAACCTCATGACCTCAGG
 TGATCCACCCACCTTGGCTCCCAAAGTGCTGGGTTACAGGCTAACGCCACCGTCCCCAGCCCCGATTCTCTTTAATT
 CATGCTGTTCTGTATGAATCTCAATCTATTGGATTTAGGTATGAGAGGATAAAATCCCACCCACTGGCAGTCAGT
 5 CAGGGAGCACCTGTGCAAGGAGCACCTGGGATAGGAGAGTTCCACCATGAGCTAACCTCTAGGTGGCTGATTTGAATG
 GCTGTGAGATTTGTCTGCAATGTTGGCTGATGAGAGTGTGAGATTGTGACAGACATTCAAGCTGGATTGATCAGTGA
 GGACGGGAGCCCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCAGGCTGGTATTAGCTCTCCGTGTCGGCCAGG
 TGACTGTGGAGGGCTTAGTCAGAAGATCAGGCTTCCCCAGCTCCCCCTGACACTCGAGTCCCTGGGGGCTTGTGAC
 ACCCCATGCCCAAATCAGGATGTCTGCAAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGCTGG
 10 GACCCGACGTGGTGTGGGCCATTCCTGATCTGGGGAGGGTCAGGGCTTCCCTGTGGAAACAAGTTAACAC
 AATGCACCTTAATTAGACTTACACGTATTAAATGGTGTGCAACATGGTCATTGACAGTATTTGGAAAGAAT
 TTAATTGGGTGACCGGAAGGAGCAGACAGACGTGGTGTGCCCCAAGATGCTCTGTCACTACTGGACTGTTCTG
 CCTGGGGGGCTTGGAGGCCCTCCCTGGACAGGGTACCGTGCCTTCTACTCTGCTGGCCTGCGGCCCTGCGTC
 AGGGCACCACTCCGGAGCACCCGGCCAGTGTCCACGGAGTGCAGGCCACAGATGCCAGGTCCAGG
 15 GTGGCCGCTCCAGCCCCGTGCCCCATGGTGGTTTGGGGAAAAGCCAAGGGCAGAGGTGTAGGGAGACTGGTGG
 CTCATGAGAGCTGATTCTGCTCTGGCTGAGCTGCCCTGAGCAGCCTCTCCGCCCTCCATCTGAAGGGATGTGGCT
 CTTTCTACCTGGGGTCTCGCTGGGCCAGCCTGGCTACCCAGTGGCTGTACCAAGGGACAGGCATCTGTG
 AGGGCATTGGGTCACGTGGCCCGAGATGCAAGCCTGGGACCAGGCTCCCTGGTGTAGGGGGACAGTCACCCGG
 20 GTTGACCGCCGACTGGCTCCCCAGGGTTGACTATAGGACCAAGGTGTCCAGGTGCCCTGCAAGTAGAGGGCTCTCAG
 AGGCGTCTGGCTGGCATGGTGGACGTGGCCGGCATGGCTTCAGCGTGTGCTGCCGTGGTGCCTGAGCCCTCAC
 TGAGTCGGTGGGGCTTGTGGCTTCCGTGAGCTTCCCCCTAGTCTGTCTGGCTGAGCAAGCCTCTGAGGGGCTCT
 CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTCTTGGTTAACCTCTTTAACAGAACGTGCGTTGAGCCCCACATTGGTATCAGCTTAGATGAAGGG
 CCCGGAGGAGGGCCACGGACACAGCCAGGGCATGGCACGGGCCAACCCATTGTGCGCACAGTGAGGTGGCCGAGG
 TGCCGGTGCCTCCAGAAACAGCGTGGGTGAGGGGAGCTCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG
 CAGCCGGCCAGGGCTGGATGCAAGCACGGCCAGGGCTCTGGATCCGTGTCTGCTGGTGCAGCCTCCGTGCGCT
 TCCGCTTACGGGGCCGGGACAGGCCAGTGCAGGAGGCCACGGGCTCTGAGGATCCTGACCTTGGCCACCG
 30 CTCCCTGCACCCACCCCTGTGGCTGGCTGGCTGACCCGTACATCTGAGGAGAGTGTGGGGTGAAGTGGACAGAG
 GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCGTTCAACAGGGTCTGAGGAAGCTGGAGGG
 TTCTAGGTCCGGGTCTGGTGGCTGGGACACTGGGAGGGCTGCTCTCCCTGGTCCCTATGGTGGGGTGGCAG
 TTGGCCGGATCCACTTCTGACTGTCTCCATGCTGCCCCGAG

Intron 5 (SEQ ID NO 8)

35 GTGGGTGCCGGGACCCCGTGAGCAGCCCTGCTGGACCTGGAGTGGCTGCCGTATTGGCACCTCATGTTGGTGGAG
 GAGGTACTCTGGTGGCCGAGGGAGTGCAGGTGACCCGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
 CTTCAGCCTTCCGTGAGCACATGGGGCCAGTGTGCACCCGTACTGCCGGGCTCTATTCCCAAGGAGGGTCCACTG
 GATTCCAGTTCCGTAGAGAACCGAACCGCTCAGCCACCGGCCCCGGTGCCTGACCCAGTCAGGCC
 40 GGGTCTCCTGTCCTGAGGCTCAGAGAGGGACACAGCCGCCCTGCCCTGGGTCTGGAGTGGTGGGGTCAAGAGAG

AGTGGGGGACACGCCAGGCCAGGCCAGGCCAGGCCAGAGGTGATGTCAGTTCTGGTGCCACTGTCAGTCCTCGC
CTCCACTCACACAG

5'-Bereich Intron 6 (SEQ ID NO 9)

5 GTAAGGTTCACTGTGATAGTCGTGTCAGGATGTGTCAGGGATATGAATGTCAGTCGTGTCAG
ATGCGTTCTGTGGTGGAGGTACTTCATGATTACACATCTGTGATATCGGTGTCGGACGTGTCATGGTGCAT
GTATCTGTGGCGTGCAATTGTGGTGTGTGTGGACGTGTCATGGTGTGTCATGGTGTGTCAG
TGTGTGTCTGTGACACGTGCATGTTCATGCTGTGTCATGTCATGTGATGTGCATATTGTGGTGTGTCAT
GTGTCCTGTGACATATCGGTGTATGGCATGGGTGTGTGGGCCCTTGGCCTACTCCTCCTCCAGGCATGGTCC
10 GCACCATTGTCTCACGCTCTGGGTGCTGGTTGGGAGCTCACATTGAGGTCTCAGGGTCTCACTCTAGCATGGTGCCT
GTCTGTGCAAGGGCTGGGCCCTGGAGACTGTAAGCCAGGTTGAGAGGAGAGTAGGGATGCTGGTGGTACCTTCTGGA
CCCCCTGGCACCCCAAGGACCCCAGTCTGGCTATGCCGCTCCATGAGATATAGGAAGGCTGATTCAAGGCTCGCTCCCC
GGGACACACTCTCCAGAGCGCCGGGGGCTGGGCTGGCAGGGTGAAGGGCCCTGGCTGGGTTCCACCC
AGTGGTCATGAGCACGCTGGAGGGTAAGCCCTAAAGTCGTGCCAGGCCGGGTGAGAGGTGAAGAAGTATCCCTGGA
15 GCTTCGGTCTGGGAGAGGCACATGTGGAAACCCACAAGGACCTCTTCTGACTTCTGAGCT

3'-Bereich Intron 6 (SEQ ID NO 10)

20 TGTGGGATTGGTTTCATGTGTCGGATAGGTGGGATCTGTGGATTGGTTTATGAGTGGGTAACACAGAGTTCAAG
GCGAGCTTCTCCTGTAGTGGGCTGCAGGTGCTCAAACAGCTTATTGAGGAGACCATATCTCCTTGAACTATGGT
CGGGTTTATAGTAAGTCAGGGGTGGAGGCCCTCCCTGGCTCCCTGTCTGTTCTTCCACTCTGGGTCGTGTTG
CCTGCTGTGGTGTGGCCGGTGGCAGGGCTTCCAGGCTCCTGTGTCATGGCTGGATGTGCCCTGGCTACGCT
CCGTCTTGGAAATTCCCTCGCAGTTGGAGGCTTCTTCTTCTTTCTTTCTTTCTTTGATAACAGA
GTCTCGCTCTTTTGCCCAGGCTGGAGTGGTTGGCGTGATCTGGCTACTGCAACCTGTGCTTCCGTAGTTCAAGCA
ATTCTCTGGCTCAGCCTCCAAGTAGCTGAATTAGGCAGGCCACACCATGCTGACTAATTGGTAAATTAGTAG
25 AGACGAGGTTCTCATGTTGCCAGGCTGGCTCGAACCTCTGACCTCAGGTGATCCTCCCACCTGGCCTCCAAAGT
GCTGGGATGACAGGTGTGAACCGCCGCGCCGGAGACTCGCTCTGAGCTCCGTGAGATCTGAGCGATAGCTG
CCTGCAGCCTGGTGTGACAACCTCCGGTTCTCCAGGCTCGCTAGGGCTTCCATTCTGACTCTTCA
CAGAAGAGTTCACGTGTGCTGATTCCGGCTGTTCTGCTGAATTGGTGTCTGTTATGATGGCTCC
TTTCTTTAGGCTTGTATTGTTCTGGCTCTGAAGGAAAAGTTGCTTATGGATGTTGAACCTTCTT
30 TCTAAACAACGATCTGAAGTTGCCCTTCCCTAAAGCAGGGATCCGAGGCCCTGGCTGGAGTGGCACGGCT
GGGGCTGTTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGTTGGGAGGCCAGCGTCCCGCCTGAGCCCCCCCC
TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACACCCACTGAGAACTGTGCGTGAGAGGGTCTAGATTCT
GTGCTCTTATGGAATCTAATGCCATGATCTGAGGTGGACCGTTGCTCCAAAACCATCCCCCTCCACTGCTG
TCTCTGGAAAAATCGTCTCCAGAAACAGTCCCTGTAACACATGGTGGGACCCCTGTGCTAAAGACCTGCTCA
35 GCAGCCTCTGTCAGTGTGATATATTGGCTTCTGTGTTGAGTCCAGAATAATTACGGATTCTGATGCTTCCGC
CGACCTCAGACCCATGGCTATTGTGGCTGTTGCCCTCTGGTGGGAAGGGTGCAGGCCCATGTAACCTTCT
GTTACTGCCTCCAGGTTGGTCTCAGGGTGAATCGTACTCGATGGTTTAGCCCACGGCCCTGCCGCCAGCTCCTG
GGGGCTGGGAAACATGCTGAAGCACAGAGTCACCGTGCCTTTGATGCCCTACAAGCTGAGGCCCTGTGTC
TGTTAGTGTGTGTCACGTGCTGTCACATCTGCTTGGGACGCAGGGCTTAGCAGGTCCGTAGTAAATGACAAGC
40 GTCCCTGGGAGTCTGCAGAATAGGAGTGGGGTGCCGGTCTCTCCCGCTTCAAGACTCTTCTGCTGTGCT

5 GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGAGCTCGAGTGCACATTGTGCCACGT
 GACTGTGGATGGCAGTCGGTCACGGGGTCTGATGTGGTACTGTGGATGGCGGTTGGTCACAGGGTCTGATGTGTG
 GTGACTGTGGATGGCGGTCGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGG
 ATGGCGCTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTG
 GGGTCTGATGTGGTACTGTGGATGGCAGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 TGGTACTGTGGATGGCAGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGG
 GTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTG
 CGGTCTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 CAGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 10 GGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 TCTGATGTGGTACTGTGGATGGCGATCGGTACAGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 GATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGG
 GACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCG
 15 CGGTCTGGTCCCGGGGTCGTATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGG
 TCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 GGTGACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCG
 GACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGG
 20 GCGGCTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGG
 GTCACAGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTG
 GGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGG
 GTGGTACTGTGGATGGTGATCGGTACAGGGTCTGATGTGGTACTGTGGATGGCGGTCGTGGGGTCTGATGTGGTACTGTGG
 ACTTTGCGCTCTCGGCCCGCCCGCCCGTTCCAAACAGAACAGCTCCAGCGCTCTGGCTCATCCGCCATCG
 25 GGCTTGGCCAGTCCACCGTCTGATCGGAAGAAAAGTCCCCAGCTCTGGGGGGCAGGCCACATTGTGGCTC
 ATGCCCTCTCTCGCCGGCAG

Intron 7 (SEQ ID NO 11)

30 GTCTGGGCACTGCCCTGCAGGGTTGGCACGGACTCCCAGCAGTGGCTCTCCCTGGCAATCACTGGCTCATGACCG
 GACAGACTGTGGCCCTGGGGGCAGTGGGGGAATGAGCTGTGATGGGGCATGATGAGCTGTGCTTGGCAAAATC
 TGAGCTGGGCATGCCAGGCTCGCACAGCTGCTCATTAGCACCTGCTCACGTTGACTGCGGGCCTCTCCAGTT
 CCGCAGTGCCTTGTTCATGATTTGCTAAATGTCCTCTGCCAGTTGATCTTGAGGCCAAAGGAAAGGTG
 35 CCTTTAGGAGGGCAGGCCATGTTGAGCCGTCTGCCAGCTGGCCCTCAGTGTGGTCTGAGGCCAAAGGAAACG
 TGTCCCCCTTCTTAGGAGGACGGGCGTGTGTTGAGCCACGCCCCGCTGAGCGGGCCTCTCAGTGTGGTCTG
 GGCCCTGTGGCCCTTGCAGATGTGGTCTGTCACGTGGCCCTGTGGCTTTGCAGATGCCCTGTTAGCA
 TCTAGGGGACAGTGTGTCACCGCATGAGGCTCAGAGACCTCTGGCGAATTCCCTGGCTCCAGGGTGGGGTGGAG
 40 GTGGCCTGGCTGCTGGGACCCAGACCTGTGCCGGCAGCTGGCAGCACTCCTGGATCACATATGCC
 CGGTGGGCTGTGGGTGAGCCAGCTGGACCCAGGGGCCAGAGGAGACGTTCTGTGTCACACACTCTGCC
 GCCCATGTGTGCTGCAGAGACTCGGCCGCCAGCCACGATGCCCTGCATTCCAGCCCAGCCCCGACTTC
 AACACTGACCCAAAAGGGACGGAGGGCTTGGCACGTGGTCTGCCCTGCTCAGCACCACGGCTCA
 TCTCCCGTCTGCTTGCAG

Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTCGGGTGGCTGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCCTGCC
 CTTCCCCACTGNCCTCTGCCGGGCCACCAGAGTCTCCTTTCTGCCCTCCGGCTCTGGGCTGGAAATGGCAAGC
 5 TCCCGAGGCCCGAACATGGCTCGGCTTGCGGAGCCGGAGCGAGGGCGGGGGTGTGCTGGTCAAGGTGTCGCCAGCGT
 GGGGTGTTGGAGTTGCTCTCGCTGGAGGACGAGGGCGGGGGTGTGCTGGTCAAGGTGTCGCCAGCGT
 GCAGCTTGTCAAGCTCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGATCTCTCTCCC
 GATAAAA
 GGATTTATCCGATTCTCATTCCTGTCTGTGACCCCCGGAGGGCGGCCCTTCTCTGTGACTAGATT
 CCCATCTGAAAGTGCAGGGTTGACCGTGTAGTTGCTCTCTGGGGGCTGTGGTGGCCATGGGCAGGCCCTGG
 10 GAGAGCTGCCGTACACAGCCACTGGTGAGCCACACTCACGGTGGTAGAGGCCACAGTGCCTGGTGC
 CCTCTGGATTTAAGTAAAACACACACCTCCCGCAGGCATCTGCCCTGCCACCCCTGTGTTGCTGG
 GGAGGAATTCTGTGCACACTCAAGGTCAAGGTCAAGGCATCCGCAGGTGGAACTGGAGGCCCTCTCTGG
 GATC
 GTCTCCAGCGATAAAGGACTGTGACAGCTTCAAAGCTTAAATATAACTATTAAATTGCATTATAAGT
 AATCACTAATGGTATCAGCAATTATAATTAAAGTATAATTAGAAATATTAGTACTACACAGTTCTGGAAAA
 15 CACAAATTGCACATGGCAGCAGAGTGAATTGGCCGAGGGACACGTGTGCACATGTGTAAGCGGCC
 AGAACATTGGCTGACAAAGTCACCTCCCCAGAGAACGCCACAGGGCCCTCTCGTGGTGTGAATT
 TAAGTGGATC
 AAGTCACGTACCGTCCACGTGTGGCAGGGCTTGGGGAAATGTGAGGTGATGACTGCGTC
 CTGCGTCCATGCCCTGACAGACAGGA
 GGTGACTGTGCTGTCTGTCCCTAGGACACGGCAGGGCCGAAGCTCTAGTCCC
 ATCGTGGTCCAGTTGGCCTCTGA
 ATAAAAACGTCTCAAAACCTGTTGCCCAAAACTAAGAACAGAGAGAGTT
 CCCATCCATGTGCTCACAGGGCGTA
 20 TCTGCTTGCCTTGACTCGCTGGCTGGCCGACTCTAGAGTTGGTGGTGTGCTCTGTG
 CAAAAGTGCAGTCCTCTT
 GCCCATCACTGTGATATCTGACCCAGCAAGGAAGCCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
 CTGTTGCTGCTGGCTTGTGAGTGCAGTGGCGCGATCTA
 ACTCACTGCAACCTCCGCTCCGGTTCCAGCATTCTC
 CTGCTCAGCCCTCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCC
 CTGGCTAATTGTGATTAGAG
 GGGTTTTGCATGTTGGCAGGCTGGTCTCGAACCTCTGACCTCAGGTGAT
 CCACCTAGCCTCCGCTCCAAAGTGC
 25 GGATTACAGGTGTGAGCCATCACGCCAGCGGAAAGCCTTTTAAGGTGACC
 ACCTATAGCGCTCCGAAATAAC
 AGGTCTGTTTGAGCTAGGCTGCAAGCGTCTCTAGCAACAGGAGTGGCG
 CCTGTTGCTGCTGGCTTGTGAGTGCAGTGGCGCGATCTA
 ACTCACTGCAACCTCCGCTCCGGTTCCAGCATTCTC
 CCCTTGCATCTGTTGGAGAGTTCTGCTCTCGTGGTCA
 CATGCTGAAACTAGGGCAAGGGTGTATCCGTTGGCG
 GCAGCGCTACATGTAGGGTCA
 CATGAGTCTTCAACCGTGGACAAATTCTTGA
 AAAAGGGAGTCCGGTTAAC
 30 TCATTCCGGTCAAGTGTCTGTTCTGTGAAATTAAACTCTA
 AGATTAAAGAAACCTTAATGAAAGAAA
 ACCTTGTGATTGATT
 AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTT
 CAGCCGCCAGTGCATGGTGA
 GAGTGGGAGCAGGGATTG
 TTTGTTGAGGGTCTCATCTGGTATGTTCTGAGGTGTT
 TGCCGGCTGA
 ATGGTAGACGTGTC
 TTGTGTTGTTGAGGT
 TCTGTTGCTGTGTTGAGGT
 GAGTGTGACGC
 ATGTCAGCACATGCC
 CTGCCGCTCTCACCTGTG
 CTCTCCCG
 CCCAG

Intron 9 (SEQ ID NO 13)

GTGAGGCCTCTCTCCCCAGGGGGCTTGGTGGGGTTGATTGCTTTGATGCATT
 CAGTGTAAATTCTGGTGC
 TCTGGAGACCATGACTGCTCTGTCTTGAGGAACCAGACAAGGTTGCAG
 GCCCCCTCTGGTATGAAGCCGACGGAGGG
 TTGACAGCCTGAGGA
 ACTGCGGCTCCACGCAGGCTCTGCTCAGCGGCC
 ATGTCCAGAGGCTCAGGGCTCAG
 GAGGCGCTGCC
 CTGCATGAGCATG
 GAATTCAACACCGAGGAAGC
 ACACCAGCTCTGTC
 ACAGTCACCCAGGT
 C

CGTTAGGGTCTTGGGAGATGGGCTGGTGAGCCTGAGGCCACATCTCCAGCAGGCCCTGACAGGTGGCTGGA
 CTGGGCCCTTCAGCCATTGCCACTGCATGGCTACACCCAAGGACGCACACACCTAAATATCGTGCC
 AACCTAATGTGGTCAACTCAGCTGGTTATTGACAGCAGTTACTTTTTTAATACTTAAAGTTCTAGGGTAC
 ATGTGCACGACGTGCAGGTTAGTACATATGTATACTGTGCATGGTGTGCTGCACCCATTAACTCATCATTACA
 5 TAGGTATATCTCTAACTGCTATCCCTCCCCTCCACTCCCCATCCATGACAGGCCCTGGTGTGATGTTCCCCACCCCTG
 TGTCCAAGTGTCTCATGGTCAGTCCCACCTGTGAGTGAGAACATGTGGTGGTTCTTCCTTGCAATAGTT
 GCTCAGAGTGTGGTTCCAGCTCGTCATGTCCTACAAAGGACATGAACCTATCCTTTTATGACTGCATAGTATT
 CCGTGGTGTATATGTGCCACATTTCTTAACTCAGTCTATCATCGATGGACATTGGGTTGGTCAAGTCTTGCTACT
 GTGAATAGTGGCGCAATAAACACATCGTGCATGTCTTATAGCAGCATGATTATAATCCTTGGGTATATACCG
 10 TAATGGGATGGCTGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCAACTGTCTCCACAATGGTGAA
 CTAGTTTACACTCCCACCAACAGTGTAAAAGTGTCTGGTGTGGAGGGATGTGGACAGCAGTTATTTTTATGAAAA
 TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTCTCGAAC
 TCCGGGTTTCTGTGCATCTTGAACTCTAGCTCAATTATAGCATGTACAGTGGATCAAGGTTCTTCATTAA
 15 GGTTCAAGTTCTAGATTGAAATAAGTTATGTAACAGAAAACAAAATTCTTGACACACAACATTGCTCTGGGATTGG
 GGAAAGTGTCTCGAGCTGGCGCACACTGGTCAGGCCCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
 GCTTGGCCTGAGGGTCACACAGTGCACCATGCCAGCTCTGTGGATAGGATCTGGTCTGGATCATGCTGAGGACC
 ACAGCTGCCATGCTGGTAAAGGGCACCACGTGGTCAGAGGGGCGAGGTTCCAGGCCAGCTTCTACCGTCTTCAG
 TTATTTTCCCTAAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCTTCTCGTCTCGACTGGCACAGAATTGACAA
 GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTACAGTTAACTGTAGAGAGCTGTCTGGAAA
 20 GAAATTAAAGTTTCTATTAAACGCTTGGAGAATGTTACTTTATTATGGCTGTGAAATTGTTGACATTCACTCC
 TCGTAGACAGATACTACGTAAGGTTAACCTTGTGTTAGTTCTTACCGTCTGGGATTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCCGTCCGTGTCGTGGGACCTCCACAGCCTGTGGCTTGAGTTGAGCCCCCGTGTCTGCCCTGG
 25 CACCGCAGCGTTGTCTGCCAAGTCCTCTCTCTGCCGGTGTGGATCCGCAAGAGCAGAGGCCCTGGCCTGACCC
 CAGGCCCTGGGGCGCACGGGCACCTCGGGAGGGAGTGGTACCGTGCAGGCCCTGGTCCTGCAGAGACGCCAGGTT
 ACACACGTGGTGAGTGCAGGCCGTGACCTGGCTCTGTGCTCTTGGAAAGTCAGAGTGGCGCTCTGGGGCCAG
 TGAGACCCCCAGGAGCTGTGACAGGCCGTGAGGGCCAGGCCAGCCTCCCTCCAGGGTGCACCTGAGCCCTGCC
 GAGCAGGAGCTGCTGAGTGAAGCTGGCCACAGCGTCTCGCTCGGTACGTTCTCGTGGGGTTGTTGGATCGTGG
 30 AGAATTGGATTGCTGAGTGTCTGCTGTAACACGGAGATGGCTAGGAGTGGCTAGAGTTGATTTTGAA
 CAAACTAAAATCAGGCACAGGGACCTGCCCTGAGCACAGGGATTGTCATGTGGCCCCCTCAAGGGGCCAACAG
 AGCCGGTGGGCTGTTAAAGTGTGCAAGGGACAGAGAACCTTGAAAGCTGAAAGGAAACCTCAGAAAATG
 TGCCGCCAGGGTGGTTTCAAGGTGCTTGTGGCTGTGTTGTGAAACCCATTGGACCCCTCAAGTCACCC
 TCCAGGTCCACCCCTCCAGGGCCGCGCTGGCTGGGGTATGCCCTGGCTCTGTGCCGCAGCCCCGAGCACAGCAGG
 35 TGTGCACATTAAATCCACTAAGATTCACTGGGGGAGCCAGGTCCAAGCACTGAGGGCTCAGGGAGTCTGAGGCT
 GCTGAGGGGAGAGCAGACGGGAAACGGTCTGTGTGGCAAGTCTGAGGGTGTGGCCAGGGAGGTGGCTCAGA
 GTGTATGTTGGGCTCCACGGGGCAGAACACTGTCTGTGAGTGGCAGCCATGTAACAGGAAGGGTGGCACAG
 GGAGCTGGGAATGCACCAAGGGAGCTGCGCAGCTGGCCAGGTCCCAGGGCAGGCCACAGGAAGGGCAGGGGACGCC
 GGGGCCACAGCAGAGGCCGAGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGCTCCCTG
 40 AGCTGGGTGAGCGAGGCTCATGACTCGCGAGGAACCTCCTTGACGTGAAGCTGACCACTGGTGTGCCAGCTCACAG

CCCAGCCAGGTCCCCGCGCTGAGCAGGAACCTCAGAACCCCTCCCTTGCTAAAGCACAGCAGATGCCCTCAGGGCATCT
 AGGAGAAAACAGGCAAAGTCGTTGAGAACAGTCTTAAAGAAGGTGGATGGTGCATGGGACATTTCTGTCCAGATTAGTCT
 GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGCATGGGACACATGAGATGGACCATCACAGAGGCCAC
 TGGGGCTGCACCTCCCCTCTGAGTCCTGGTGTCCAGGGCAGGTTCTGCATGCTCACCTACCTGCTGC
 5 GGGAGACAGGAAAGCACCCCAAGTCTGGAGCAGGGCTGGGTCAGGGCTCTCAGAGCTCTGCCAGGCCAGCACCC
 GCTCCAATCACCCTCTGGGTTTCCAAAGCATTAAACAAGGTGTCAAGTTACCTCCTGGGTGACGCCCGCA
 TCCTGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAAGTGGAGCCTGTGCCCGCTGGGCAGGTGCTGCTGCAGGGCGTTGGTCCACCTCTGCT
 TCCGTGCCCCAGGGCACTGCCAATCCAAAGGGTCAAGGGCCACAGGGTCCCCCTGCTCCATCTGGGCTGAGCAGA
 AATGCATCTTCTGTGGAGTGAGGTGCTCACAAACGGGAGCAGTTCTGTCTATTGGTAAAGGAAATGGTGCAC
 CAGACCTGGGTGCACTGAGGTGCTCAGAAAGCAGTCTGGATCCGAACCAAGACGCCGGGCGCTGTGGCGTGA
 CTCTCAAACCGAACACAGGGCCCTGCTGGCATGAGTCCCTCTGAACCCGAGACCTGGGCGCTGTGGCGTGA
 15 CTCTCCGAACCCAGAGACTCAGGGCCCTTGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC
 AGTCTACAGGATGCCATGAGTCATGATCACGTGTGACCCATCAGGGACAGGCCATGGTGTGGGGGGCTCTACAA
 AATTCTGGGTCTGTGTTCCCCAGAGGCCAGAGCTCAAGGCCGCTCAGGCTCAGACACAAATGAATTGAAGATGGA
 CACAGATGCAGAAATCTGTGCTGTTCTTTATGAATAAAAGTATCAACATTCCAGCAGGCCAGGTGGCTCACACCT
 ATAATCCAGCACTTGGGAGGCCAGGTGGGATCACTTGAGGCCAGGAGTTGAGGCCACCTAACCAACATAGTG
 20 AAATTCCATTCTACTTAAAAAAATACAAAATTAGCCTGGCTGGACACGCCCTGTAGTCCCCCTATGCCGGAGGC
 TGAGGCAGGAGAATCTTGAACCCAGGAGGAGGGTTCAGTGAGCCAGATCACACCACTGCACTCCAGCCTGGCA
 ACAGAGTGAAGACTTCATCTTAAAAAAAGTATCAGCATTCAAACATAGTGGACAGGTGGCTCACACCT
 TGTCCTCGATAATATTACTGGTCTGTCTAGAGGCCAGACTGGGGTGCCCTCTGAAAGGCACACCTCATGG
 GAAGAGAAATAAGTGGTAATGGTTGTTAAACAGAGGTTAACTGGGTCTGTCTAGGTTAACAGTCCAGATC
 25 TGGACTTTGCCCTTCCAGAATGCTCCCTGGGTTGGCTCATGGGGAGCAGCAGGTGGACACCCCTGATGGG
 GAGCAGCAGGTGCAAGCCTCATGATGGGGAGTGGCAGGTGCAAGACACCCCTGTCATGGGCCAGCATGCTCG
 TTGCACTCCCTCCCCACAAGGATGCCGTCTCTGTGCTCCACAGTCCCTGCTTCCCTCTCACAGCCTAACCTGG
 CTGGCCCTCCACTGGCTTGTCTGATGATTCCACATTCTGGCTCCAGCACCTCTGCCCTCTCCAGGCACCT
 GCAGTGTGGCCATACCAGTCAGCTGTGAACTGTCCACTGCTTATTTCTGCCCTGAAATGTTAGGACAGGC
 30 ACCCCTGTTCCAGCCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAGGACAGACAAACAAATCAGAAAATGG
 GTTCTCTAAACACATTGCAAGGCCACAGAGGCTAGTGCAAGGATGGGGCATCAGGTCTAGTGGTGGCTAACATG
 CCAGAATATTCTGTGCTCCAAAGGCCACTGGTCAGAGTGTGCTGAGGTGGCTCTAAAGCTCAGCAGTGGAG
 GCAGTGGTGCCTACTCAGGGTAACATCCTCTGTGCTGAAATACAGCAGAGGCTTAAGGGCATCTGGGA
 GAAGAAAACAGGCAAATGATTAAGAAAAGTGGTAAGATGGGAATTCTGTCCAGATTAGTC
 35 TCCCCAAACACAGCTCAGATGGTAGAATGTGGTCAGAACATGATGGACAGAACATAGAACAAACGGAAAGCCCTATCT
 CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTGTAATTCTGAGAAA
 GACTGGGAAGCAATAAGTTGTCTTACAGCATATACAGCAGAGCAGATTCTAGGTAGAAGAGGAGACACATGCA
 ACCAGCAACAGAAATAAAACAAAGACTCAAAGGAAGGGAGGTGAACTGGCTGGTTGGTGTGGGAAGGACACAC
 AGGGAGGCGGATGAAACCAGTGAGGCCACGGGCTTGTGCTTCACTGCAGAGAAACTCAGCTGCCAGCCACAGTGA
 40 ATGGCCATTCCCTGGAGCGTTGTGCACGTGATTAAAGGCCCTGTGAGGTCTGCACATTCACTCTCACTT

5 GTTCTCTAACCAACCTGAGAGGTAGAGGGAGGAAGGCTCAGGGAGCAGCCGCCCTGGTCACCCAGCTGGCAAAGGGC
 ATGCATGATTGCAGCCCTGGCCTCTGCTCCGGGCCCTGCTCTGCCAGGACCCACACAAGTCAGACCCATAGGCTC
 AGGGTAGGCCGGAGCCAAAGGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
 CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC
 GAGACCTGCCCCATCCCTCATGCTGGCTCTTCTGGGCTTGCCAAGAGCCAGCATCAGTTGAGGCAAGCTGGAAAG
 ACTTTCTGAAAGCAGCTGTTGCATGGAAGTCTCACAAATGTCCTGTCTTCCCAGTAATTCACTTCTGAAGTGA
 CCAGACATTATCACGGGCTTATTACATTCCAGTGTCCAGGGCAGGGGACTGCCACAGCAAGTCACGAACCTGCC
 CAAATAACAGGGCTAAGGAGATAATTATGCATCACAAACTTGTCTGCCATTAAACATTTCAAAGAATTGGAAAGAAT
 GTTTAATGGCACAAAACGTTTATTCAATGTAGCAGTGTCAAAGCTGGATGTAAGAACACACCCAGGAGCCTGCC
 10 TGAATGTCATGTGTTCATCTTGACATGGACATACATGGCAGTGTGAGTGGTGTGAGGACATCGGTGG
 GATGCCCTCCATCCCTGGCCCTCTGGAGACACCATGGTGCACAGTGGCAGTGTGACTCACTGGAGCCCTGTTAGCTGGTGCACCTG
 GCTCTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCAACTCAGTGTCTCCACAAAAACCTGAGTCAC
 ACCTGTGTTCACTCGAGGGACGCCGGCAGGGCTCCACAGTTATTATGTGTTTGCTGAGTTATGTGCAAGATC
 TCATCAGGGCAGATGATGAGTGACAAACACGGCGTGGCAGGGTTGGATAACACTCAACATCACTAGCCAGGTCTGGTG
 15 GAGTTGGTCATGCAAGAGTCGGATGGCATGTAGCATTGGAGTCCATGGAGTGAGCACCCAGGCCCTGGCTGCAGC
 GCATGCCCAAGGAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTTTGGAGCAAGCTTGCAGGAGGGGCTGGGTG
 GGGCAGGCACCTGTGTCACATTCCCCCTGTGTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTAGCAGAGTCAGGAGGTGTGCGCAAGTATGTGTTGTCGCGCGCT
 GCCTGCAAGGCTGATGGTCACTGGCTGCAGCTAAGAGTCAGCATGTACGCATATACACGTGAGCACATACATGTGTCAT
 GTGTCATGTCATGAAGGCATGGCAGTGTGTCACAGGTGCAAGGGCACAAGTGTGTCACATGCAATGCAACACCTGACA
 TGCATGTGTCGTGTCACAGTCGTGTGGCATTACAGTGAGGTGCATCGTGTGGGTGTGAGTGTGAGTAGCATGTG
 GCACATAACATGATTGAGGGCTCTGTGTTCACCCCTAGGTCTCAGCACCAGTGCACCTCTTACAGGATGAGAC
 25 GGGGTCCCAGGCCTTGGTGGCTGAGGCTCTGAAGCTGAGCCCTGAGGGCATGGTCCATGGGATCCGTCACCTCT
 CCCCTCTGTGGCTTGTGTCACCTCCCCCTCTGGGCAATTACATCCACTCCACTCCCTCTCTGTGG
 ATCCCGTCCACTCCCCCTCTGTGGGATCTGGTCCACCTCCCCCTCTGTGGGCAATTACATCCACTCCACTCCCTCT
 GGTTCTTCTGTCTGGCCGAGCCTGGGGCAGGAGATGACACAGAGTCGGACTCGCCAGGGTGGTTCGAGCTG
 CCGGGTGAGGCCAGGGCGATTCACTGGGAAGAGGGATAGTTCTGTCAAATGTTCTCTTGTGTCATCTGA
 30 ATGGATGATAAGCAAAAGTAAAACATTAAACATCCAGAGAGGTTCTACCGTTCTACACTTCTGGCAGCTAG

Intron 13 (SEQ ID NO 17)

GTGAGCCACCAAGGGTGCAGGCCAGCCTCCAGGGACCCCTCCGGCTCTGTCACCTCTGACCCGGGCTTCACCT
 35 TGGAACTCTGGTTTAGGGCAAGGAATGTCCTACGTTCTAGTGGTGTGCTGCTGTGACAGTCTGTTCCGCTG
 GCTCTGTGCAAAGCACCTGTTCTCATCTGGTAGTGGTAGGAGCCGGTGTGGCCCAAGGTGCCCCACTGTGCTGT
 GCACCTGGCCCTGGGACGTATGGAGGCCATCCAGGGCAGAGGGCATGGGAAAGAGATGTTATGGGAGTCTTAC
 CAGAGGAGGCTGGGAAGGTGTCGAACAGTAGATGGGAGATCAGATGCCGGAGGATTTGGGTCTCAGCAAAGAGGG
 GAGGTGGGTGAGGTGAGGGTGCCTGGCCCCACCCCGGGAGGTGCAAGCAGAGCTGGCTCCCAACAGCCGGCCA
 GCACCTGTGTCCTGGGATGGCTGTGCTCTGGAACGTTCCCTGTCTGGCTGGTCAAGGGGTGCCCAAGAATCG
 40 ACAACTTATCACAGAGGGAAAGGCCAATCTGTGGAGGCCAGGGCAGCTCTGCCAGTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTGGGGCTGTACCAAAGGGCACTCGGCACACAGGCCGGGCTCCACCTAACAGGCCTCCGAGCCACTG
 GGAGCTGAATGCCAGGAGGCCAAGGCCCTGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTTACCCAGGGCG
 AGGCTGCGGAATTACCGTGACACTTGATGTGAAATGAGGCTGTCGTATCGTGGAAACCCAGCAAGGGCTCACGGGA
 GAGTTTCCATTACAAGGTGTCACCATGAAAATGGTTTAACCGAGTGTGCGCCCTCATGCTCTGGCAGGGAGGGC
 5 AGAGCCACAGCTGCATGTTACCCCTTGGCACAGCTCAGGGCTTGGGACCAGGCTGTCAGTCCAGGGTGCCTCC
 GGCTCAGACCGCCCTCTCTGCTCTCTGCCTCAAATCTCCCTGTTGCATCTCCCTGACGGGTGCCTGG
 CCCTCGTCAAGCTGCTTGACTCTTCGGAAACCCCTGGGTGTCGGATAACGGTGCCACTGAGGACTGGAGGTG
 CTGACACTGTGTTGACCCCAGGGTCCAGTGGCGTGCTGGGCTCTGGCATGATGAGGTCAAGAGGAGTTTCC
 CAGGTGAAAACCTGGAAACTCCCAGGGCATGTGACCTGCCACCTGCTCTCCATATTAGCTCAGTCTTGCTCTC
 10 ATTTCCCCACCAAGGGTCTAGCTCCGAGGAGCTCCGTAGAGGGCTGGGCTCAGGGCAGGGGGCTGAGTTCCCCAC
 CCATGTGGGACCCCTGGTAGCTGCTGATTGGTAGCCCTGAGGAGGCCAGATGGCATGGGCCACGGGCCGTTTCCA
 AACACAGAGTCAGGCAGCTGGAAGGCCAGGAATCCCTCCCTGAGGAGGAGTGGAGAACGGAGAGCTGGCCCCG
 ATTTACGGCAGCCAGGCTGCAGTGGCGAGGCTGTGGTGTCCACGTGGCGCTGGGGGGGTCTGATTCAAATCCGC
 TGGGCTCGCCTTCCGGCCCGTGCTGGCGCGCTCACGGGTTGGACGCCACCTCTAGCAGGTGGC
 15 TATTCCTCCCTTGGAAAGAGAGGCCCTCACCATGCTAGGTGTTCCCTCTGGGTAGGGACGGAGCTGGCGTGGCAACC
 CGGGGACCTTGGCTTATTTATTTGTTAAAAACATTCTGGCCTGGCTTGGCTAAATGGGAAAAGACATCC
 CACCTCAGCAGAGTTACTGAGAGGCTGAAACGGGGTGTGGCTTGTGACTGGTGTGATCTCAGGTCAATTCAAAGTGGCT
 CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGTGAGATGAGGTACACGGGGCTCAGGCAGTGGGT
 GAGGCCAGGTACATGGGGGCTCAGGCAGTGGTGAGATGAGGTACACGGGGCTCAGGCAGAGGGTCAGACCAGGTAC
 20 ACGGGGCTGATCACACGACATATGAGCACATGTGACATGTGCTGTTCATGGTAGCCAGGTCTGTGACACCTGC
 CCCAAAGTCCAGGAAGCTGAGAGGCCAAAGATGGGGCTGACAGGGCTGGCGCGTGCACACCTGAGTCCAGCA
 CTTGGGAGGCCAGGCAGAGGATCCCTTGAGCCAGGAGTTAACGACAGCCCTGAGCAACATAGTAGAACCCCCATCTC
 TATGAAAAAATAAAACAAAAATTAGCTGAACATGGGGTGTGCGCTGTAGTTCAAATACTTGGAGGCTGAAGTGGGAG
 GATCACTTGAGGCCAGGGTGGAGCTGAGCTGAGATGGCACACTGTACTGCAGGCTGGTGACAGAGTGGAGA
 25 GCCCACATCTAACACAACAAAGAAAGACTGACAAATGCAAGTTCTGGAAAGAAACATTAGTAGGAACCTAACCTACACA
 CAGAAGCCAAGTCGGTGTCTGGTGCACTGAGATGAGATGAGGTACGGTCTCACACCATCCCCAGACCCAGGGTTATG
 CACCAACAGGGGGGGTGGCTCAGAGGGATGCGCAGGACGGTGTGATATACGATGACATCAAGGGTGTGACGAAGGGCAG
 GATTGATGATAAGTACCTGCTGGTACACAGGAACATGGATAAAACTGGAAACCTTAGAGGCCCTCCCGAACAGGGCT
 AATCAGAACGCAGCATGGGGGCTGGCATCAGGATGGAGCTGCTCAGGCTCCACATGCGTGTACAGATGGTCA
 30 CAGAAACGCAGTGTACCTGTCACACAGACACGCACTGCACACACAAGCACACACAGACATGCATGCATGC
 ATCCGTGTGTGACCTGTCACCTGCCCCATGAGGAACCCATGCACTGCACTCATGCAACGCACACAGGACCCGGTGGCCAT
 GCCCACACCCACGAGCACCGTCTGATTAGGAGGCCCTCTGACGCTGTCGCCATCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGCAGGTGCCCTGGCTCAGTGGCAGCAGTGCCTGCCGTGGTTAGTGTGTCAGGAGACTGAGTGAATCTGGG
 CTTAGGAAGTTCTTACCCCTTTGCACTCAGGAAGTGGTTAACCAACCAACTGTCAGGCTCGTCTGCCGCCCTCTCGT
 GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGTCGGAGCTGCCATCTCCACCTTGTCTGCCCTGGGAA
 GCGCTGGGGGCCCTGGTCTCTCTGTTGCCCATGGTGGATTGGGGCTGCCCTCCTGTTGCCCTGTGGTGG
 GATTGGGCTGTCCTCCGTCATGGCACTTAGGCCCTTGTGCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG
 40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGTATCACACAGAGGCCCGCCTCTGCTTCCAGTCACCG

TCCCTGCCCCGGACACTTGTCCAGCATCAGGGAGGTTCTGATCCGCTGAAATTCAAGCCATGTCGAACCTGCGGT
 CCTGAGCTAACAGCTTCACTTTCTGTTCTGTGTTGAAATTTCACCTGAGAAGCCAGAAAACATTCTG
 TCGTGAACCTCGCGGTGCTGGGTGGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGAGCTTCCG
 GTGTCTCCTGGGAGGGAGCTGGCTGGCCTGTGACTCCTCAGCCTCTGTTTCCCCAG

5

Intron 15 (SEQ ID NO 19)

GCAAGTGTGGTGGAGGCCAGTGCGGGCCCACCTGCCAGGGTCATCCTGAACCCCTGTGTGGGGGAGCACGCC
 AGATGCTGCTGAAGTGCAGACGCCCGGGCTGACCCCTGGGACGCCACGCTGGCAGCCATGTGATTAACG
 CTGGTGTCCCCAGGCCAGGCTGGCAGGGTCCCAACTTCTGAACCCCTGCTCCATCTCAGGGCGATGGCTCC
 10 CCACGCTTGGGAGGCCCTCTGACCCCTGACCTGTGTCCTCTCACGCCCTTCCCTGGCTGTCGCCCTGAGCTCCTGGGT
 CCTGAGCAAGTCTCTCCCCGCCCCGCGCTCCAGCGTCACTGGCTGCCTGTCTGTCGCCCGGTGGAGGGTGTCTG
 TCCCTCACTGAGGTTCCCACCAGCCAGGCCACGAGGTGCAAGGCCCTGCTGCCGCCACCCACAGTCCTAGGAGGG
 TTGGAGGATGCCACCTCTGGCCTCTTCTGAAACGGAGTCTGATTTGGCCCCGAG

15

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTGAATCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGTGGT
 GGTCACTGGGCCCCTGGCTGTCATTTACGAAAGTCTATGAGTGAATGGGGTGTGGTCACTGGGGCCCCTG
 GCCTGGCTGGGCTGGAGGTTCTGATGCTGTGAGGCAAGGAGGGTAGGGGATAGACAGTGGAGCCCCA
 CCCTGGAAAGACATAACAGTAAGTCCAGGCCAGGGCAGCAGGGATGCTGGGGCCAGCTTGGCGGGGGATGATG
 20 GAGGGCCTGCCAGGGTGGCAGGGATGATGGGGCCCAGCTGGGTTGCAGGGTGATGGGGGGCTGGTCTGGTGG
 CGGGGAAGATGGGAAGCTGGCTGGCCCCCTCCTCCCTGCCCTCCACCTGCAGCGTGGATCCGGATGTGCTCCCT
 GGTGCACATCTCTGGCCATCAGCTTCACTGGAGGTGGGGGAGGGCATGACACCATCTGTATAAAATCCAGGATT
 CCTCTCTGAAAGCCCCACTCAGTTGAAAGTCACATTCCGCTCTGGCATTCTTAAGAGTAGACCAAGGATTCTG
 ATCTCTGAAGGGTGGTAGGGTGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGTGGGCTGGTGTGCTCTC
 25 ATCCTCTTATCATCTCCAGTCTCATCTCATCCTCTTATCATCTCCAGTCTCATCTCTTATCTCTAGTCTCATCCAGT
 CTCATCTGTCATCTCTTACCATCTCCAGTCTCATCTCTTATCTCTAGTCTCATCCAGTACCTCCAG
 GGGCGGGTGCCAGGCTCGAGTGGAGCTGGACATACGTCCTCCTCAGGCAGAAGGAACGGAAAGGATTGAGAACAG
 GAGGGGCGGCTCAGAGGGACGAGCTGGGGTGAAGAACAGGCCCTCCAGAAGTTGGCTGGGCCACAGAAACCG
 AGGGCCCTGGGTGAGTGGTCCAGAGCCTCCAGCAGGCCCTGGGGCTTATGGTATGGCCGGTCTACTGAGTG
 30 CACCTGGACAGGGCTCTGGTTGAGTGCAGCCGGACGTGCGCTGGTGTGGGGGGCTTATGGCACTGGATATG
 GCGTCATTATTGCTGCTGCTCAGAGAATGTCAGTGAACGACCTAATGTTATGGGGCCAAGTCCACAGACTG
 TGTCGAAATGCACTCTGGTGCCTGGAGGCCGCTAGGGAGCTGTGAGGAGAGGGCTTGGCAGGCCCTGGGG
 GCGCTTGGCCCTGCAAACGGAGGGAGCGGCCGGGGCGCCGTGGCGACGACCTCAAGTGAAGGGGGTGGACAGAAC
 AGGGCGGGACTTCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGGTTGAATCACAGACCAACAGTCAGGGCATT
 35 GTTCAGCTATCCATCTTACAAAGCTCCAGATTCTGTTCTCCGGGTGTTTTGAAATTACTCAGGATTACT
 TATATTTTTGCTAAAGTATTAGACCTAAAAAGGTATTGCTTGTATGGCTTAACACTCAACTAACCTACTTTAT
 TTGTCGTTTATTATTATTATTAGAGATGGTGTCACTCTGTCACCCAGGTTGTTAGTGCAGTGGCAC
 AGTCATGGCTCGCTGTAGCGCAAACCCAGGCTCAAGTGAATCTCCGGCTCAGCTCCAGAGTGTGGATTACAG
 GTGTGAGGCCAGTGGCTGGCACTTTAAAAACCACTATGTAAGGTCAAGGTCAAGGTCAGTGGCTCCACACCTGTATCC
 40 CAGTAGTTGGGAGGCCAGAGGATTGTCAGGCCAGGAGTTGAGGACCATGGGAAACATAGGGAGACCC

ATCTCTACAAAAAATGCAAAAGTTATCGGGCGTGGGTCCAGCATCTGAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGGAGGTATGGCTGCAGTGAGCTGTGATTGACCATCGCACTCCAGCCTGGCAACAGACTGA
GACCCCTGTCTCAAAAAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGAAGAAGGAAAGAGAAGAAGAAGAAG
GAAGAAGGAAGAAGAAGGAGAAGGAGGCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAATGAAAATAACA
5 AAGTTTAAAGGGAAAGAAAACCCAGCTTTGACTTCCTTAGGCCTGAACCTCATCTCAAGCAGCTTCCCTCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAACGAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTCTGCCTCAGCCCCACGCTCCTGCCGCTCTGCACCTGCTGAACCGTC
GATGTGGTGCAGGTGCCACCTGGGAAGGATGCTGTCAGGGGCTTGCCAAACTTGGTGGGTTTCAGAAGCCCCAG
GCACATTGTGGCAGGCACAATTACAGCCCCCTCCCAAAGATGCCACGTCCTCTCCCTGGAACCTGTGAATGTGTCACCCG
10 CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTATCCTGGATTATCTGG
TGGGCTGATATGCCACAAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGAGACTAGAGAGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTGAGATGGAGGGAGGGTCCCCAGCCAAGGAATGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCCGGTCCGTAGGGCACACGGCCCTGCCACGCCATGCTGATTTCAGGCCAGTGGGACCTGTTTCAGCTTC
CGGCCTCCAGACTGTAAGATGATGCGTTGTGTTCAAGCCACTAAGCTGCAGTGACTCTCAGCCCACCCCTGG
15 CAGTACAGGGAAATGAATAACAGGGACAGTTCTCAGAGTGACTCTCAGCCCACCCCTGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-
5 charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in
15 zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der
20 Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des
25 offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative,
30 3'-lokalierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, lässt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer interner 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur 10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv markierter Primer (5'GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10⁵ cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL), 2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser 15 erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden 2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150µl 20 Phenol/Cloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-Aacetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5 min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel 25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom 30 ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA_nTTGT) repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_nNa/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes
5 konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzurorden ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den
10 Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.
20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem
genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für
30 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer CSS: 5'-CTGCGTCCTGCTGCCACGTGGGAAGC-5', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGCGCGAGTTCAAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCCGCGC-3', 5'-flankierende Region -191 - - -171; Primer RP-3A: 5'-10 CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, CSS-15 C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

20 In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. 25 bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

30 Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpacket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschiedene lange hTC-Promotorsequenzabschnitte generiert und 5' vor 10 das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 15 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung 20 bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTCC-3', Position -4093 bis -4070) 25 wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaars PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGATTTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination 30 PK-3A und PK-5D (5'-

GGAGATCTGATCTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmente 5 (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.
10
15

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection 25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit 10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszuplattieren.

15 Für die Transfektion wurden pro Welt jeweils 2x 10³ -HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV
β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serumfreies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luciferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luciferase generierte Licht gemessen.

30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysat und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na₂CO₃ gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschiedene lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK 10 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im 15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 **Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B.** (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 **Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K.** (1987). Current protocols in molecular biology. Greene Publishing Associates and Whiley-Intersciences, New York.
- 15 **Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D.** (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 20 **Broccoli, D., Young, J. W. und deLange, T.** (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- 25 **Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S.** (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 30 **Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B.** (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 35 **Geng, Y., and Johnson, L.F.** (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol 14:4894.
- Goldstein, S.** (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- Harley, C.B., Futcher, A.B., Greider, C.W., 1990.** Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* **155**, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* **57**, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234-282.
- Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* **86**, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* **41**, 181-190.
- 25 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* **75**, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* **15**, 7155-7174.
- 30 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* **57**:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 **Shay, J. W. (1997).** Telomerase and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

10 **Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Lansdorp, P.M. (1994).** Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

15 **Wick, M., Härenen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995).** Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.

5

2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.

10

3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.

15

4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.

20

5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.

6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.

25

7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.

8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

30

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:

5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,

10 B. Messung des Substanzeffektes auf die Reportergenexpression.

10 10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.

15 11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.

20 12. Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

25 A. Inkubation eines rekombinanten Konstruktions bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält mit Körperflüssigkeiten oder zellulären Proben,

B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

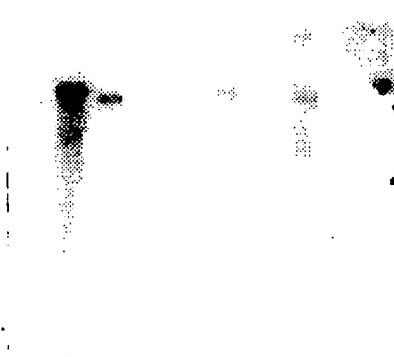
A

1 2 3 4 5 6 7 8 9 10



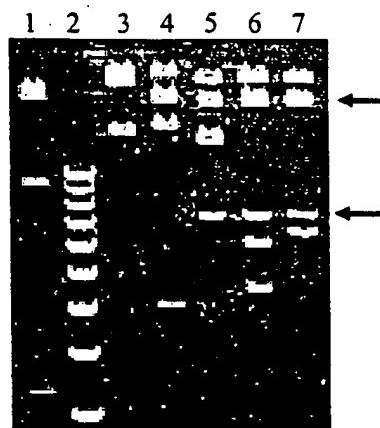
B

1 2 3 4 5 6 7 8 9 10



ERSATZBLATT (REGEL 26)

Fig. 2



ERSATZBLATT (REGEL 26)

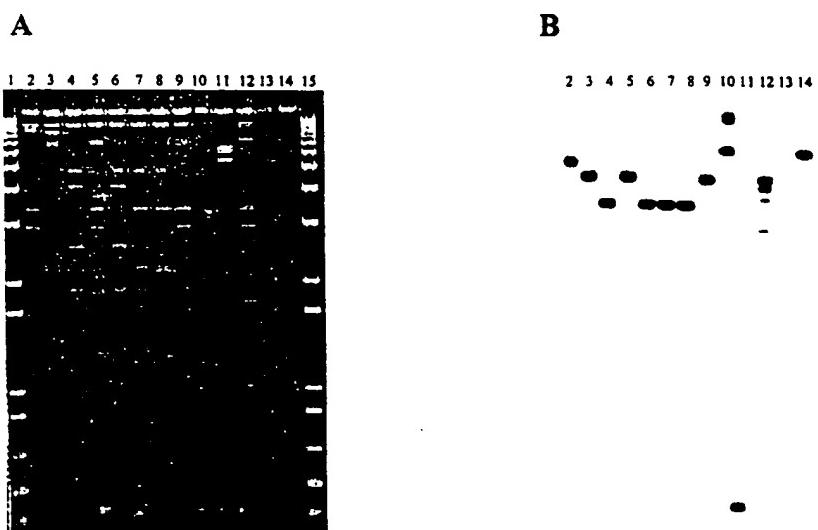
Fig. 3**ERSATZBLATT (REGEL 26)**

Fig. 4

GAGCTCTGAA CCGTGGRAAC GAACATGACC CTTGCCGCC TGCTTCCTG GGTGGGTCAA GGGTAATGAA 70
 GTGGTGTGCA GAAAATGGCC ATGTAATTAA CACGACTCTG CTGATGGGG ACGTTCCTTC CATCATTATT 140
 CATCTTACCA CCCAAGGACT GATGATTCTC AGCAACTCTC TCGGGGTGGA CAAGCCATGA CAAAACCTAG 210
 TACAACACC ACTCTTTAAC TAGGCCCACR GAGCACGGGC CACACCCCTG ATATTTAAG AGTCCAGGAG 280
 AGATGAGGCT GCTTTAGGCC ACCAGGCTGG GGTGACAACA CGGGCTGAAAC AGTCTGTTCC TCTAGACTAG 350
 TAGACCCTGG CAGGCACTCC CCCAAATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGGCCGC ATCTGCCCTG 420
 GAGACTCAGC CTGGGTGCCC ACACGTGAGGC CAGGCCCTGTC TCCACACCCCT CCGCCCTCCAG GCCTCAGCTT 490
 CTCCAGCAGC TTCTAAACC CTGGGTGGGC CGTGTTCAG CGCTACTCTC TCACATGTC CACTGTGCT 560
 TGTCTCAGCG ACCTGAGCTG CACGGCTCTC CTCACATGG GGTGTCGTC TCCCTCCCCA ACACTCACAT 630
 GCGTGAAGG GAGGGAGATT CCGCCCTCCC AGACTGCTC CTGAGGCTG GAACCTGGCT CGTGGCCCCC 700
 GATGAGGTT CCTGGCGTCC GGCTGCAAGC TGACCTCAT TTCCAGGGCC TCCCCCTCTC CTGTCATCTG 770
 CCGGGGCTG CGGGTGTGTT TTCTGTTTC TGTCGCTCTT TCCACGTCCA GCTGGGTGTC TCTCTGCCCG 840
 CTAGGGTCTC GGGGTTTTA TAGGCATAGG ACGGGGCGT GGTGGGCCAG GGGCTCTTG GGAAATGCAA 910
 CATTGGGTG TGAAGTAGG AGTGGCTGTC CTCACCTAGG TCCACGGCA CAGGGCTGGG GATGGAGGCC 980
 CCCCCAGGGG CCCGCCCTC TCTGCCAGC ACCTTCTCTC CTGGACACACA GAGTGGCAGT 1050
 TTCCACAAGC ACTAAGCATC CTCTTCCAA AAGACCCAGC ATTGGCAGCC CGGACATTT GCCCCACAGC 1120
 CCTGGAAATT CACGTGACTA CGCACATCAT GTACACRCTC CGTCCACAGA CGCACCCCCCG CTGTTTTATT 1190
 TTAATAGCTA CAAACCCAGG AAATCCCTGC TAAAATGTC TTTAACAAAC TGGTTAAACA AACGGGTCCA 1260
 TCCGACCGT GGACAGTTC TCACAGTGA GAGGAACATG CGCTTATAAG AGCTTCAGG CACTCTCAGG 1330
 GAATTACGCT GAGTCAAACAC TGGGACATTC AGTCACCAT GCTCAGGAAAG AAAGAATTTC 1400
 ACCCCATGGC AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGGCCCA GCTGGGGCT ACTGCACCA 1470
 CCTTTACTA AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CRTAGGGAG 1540
 TGGGATGGG GAAACCCGGG GGCTGTGCCA TCTTCTCCAT GCGCGAGTGT CCTGGCAGG ATAATGCTCT 1610
 AGAGATGCCG ACCTCTGTAT GTGATCTCG GAGGTCTGGG ATCCCTTGGG ACTACCTGCA GGGCCAGGG 1680
 GTGATCTCG TGAGGACCTC GAGGTCTGGG TGTGAATCTA GGATTTATTC AAAACAAAGG TTACAGAAA CATCCAGG 1750
 GGTTCTGGG AGAGGGGGC AGGAGGGTCA GAGGGGGCA CGCTCAGGAC GATGGAGGCC CTGAGCTGCA 1820
 GGCTGAAAAG GGAGGGAGGG CCTCGAGCCC AGGCCTGCA CGGCCTCCAG AAGCTGAAA AACGGGGRA 1890
 GGGACCCCTC ACAGGACCTG CACCGAGAAG GCACGGCTGG CCCTTAGGCC ACCAGGGCC ATCGTGGACC 1960
 TCCGGCTCTC GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTTGGGG ATTGCGAA 2030
 GCAACAGGAA ACCCATGAC TGTGAATCTA GGATTTATTC AAAACAAAGG TTACAGAAA CATCCAGG 2100
 CAGGGCTGAA GTGCCCTCCGG GCAAGGGCAG CGCAGGCACTG ACTGATTATG TTACTTATT TTACTTATT 2170
 TACTACTTT CTGAGACAGA GTTATGCTCT GTTGTCCCAG GCTGGAGTGC AGGGCCTGAA TCTTGGCTCA 2240
 CTGCAACCTC CGTCTCTGG GTTCAAGCAA TTCTCGTGC TCAGCTCCCA AAGTACGTTG GATTTCAGGC 2310
 GTGACCACCC ACACCCGGCT AATTTGTAT TTCTAGTGA GATGGGTTT CACCATGTTG GTCAAGGCTGA 2380
 TCTCAAAATC CTGACCTCAG GTGATCCGC CACCTCAGC TGTGAATCTA GGCTCAAGT ACSSCATGAGC 2450
 CACTGCACCT GGCTTATTTA ACCATTTAA AACTCCCTG GGCTCAAGT ACACCCACTG GTAAGGAGTT 2520
 CATGGAGTTC AATTTCCCTT TTACTCAGGA GTTACCCCTC TTGATATTG TCTGTAATTC TTCTGTAGACT 2590
 GGGGATACAC CGTCTCTGCA CATACTCACA GTTCTGTGCA CCACCTGTTA TCCCTAGGGG CCGACTGAG 2660
 GGGCAGCTGG GAGGCTGCAG GTTCTCGGTC CCAGGGGGT GCGCCATCTC CAGTAGAAAC CTGATGTAGA 2730
 ATCAAGGCGC AAGTGTGGAC ACTGTCCTGA ACTCTCAATG CTCACTGTTG GCTGAATACAT GTAGAAATT 2800
 AAGTCCATCC CTCTCTACTCT ACTGGGATTG AGCCCCCTTC CTATCCCCCC CGAGGGGGAG AGGAGTCCCT 2870
 CTCACCTCTG TGAGGAAAG AATGATACTT TGTTATTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT 2940
 TGTGCTTGT TTGTTGTGAGG GGTTCATCTC TTGTTGCTCA GGCTGGAGGG AGTGCATGG 3010
 CGCGATCTG GCTTACTGCC GCCTCTGCC CTCCAGGTTCA AGTGTCTC CTGCTCCGC CTCCCAATTG 3080
 GCTGGGATTA CAGGGCACCCG CCACCATGCC CAGCTTAATT TTGTTATTT TTGAGAGAC GGGGGTGGGT 3150
 GGGGTTCCACC ATGTTGGCCA GGCTGGTCTC GAACTCTGA CCTCAGATGA TCCACCTGCC TCTGCCCTCT 3220
 AAAGTCTGG GATTACAGT GTGAGGCCACG ATGCCCACT CAGAATTAC TCTGTTAGA AACATCTGGG 3290
 TCTGAGGTAG GAGGCTCACC CCACACTAAG GTTGTGCTGT TTTAAGGCAAR TGATGAAATT TTGTTATTTG 3360
 TGTGAGAACA CTCTGTATGT TTACACTGT GATGACTAAG ACATCTCAG CTTTCAAG ACACACTAAC 3430
 TGCACCCATA ATACTGGGGT GTCTCTCTGGG TATCAGCAAT CTTCTATTGAA TGCCCTGGAGG CGTTTCTCTG 3500
 CCATGCACAT GGTGTTAAATT ACTCCAGCAT AATCTCTGC TTCCATTCTC TCTCTCCCT TTGTTAAAT 3570
 TGTGTTCTG ATGTTGGCTT CTCTGAGAG AACCAGTGA AGCTACAACT TAACTTTGT TGGAACAAAT 3640
 TTCCCAACCC GCCCCCTTGC CCTACTGGCA GAGACAACTC ACAACACAC CCCTTAAAAA AGCTTAGGG 3710
 ATCACTAAGG GATTTCTAG AAGAGGACAC TGTAACTCTA AGTATTACA AGACAGGGCT AACCTCCAGC 3780
 GAGGGTACCA CCCAGGGAGG GGTGGGAGGC CTGTTCAATT GCTACTCTCA TAAATTAAGG AATTTCTCC 3850
 GGCAGTTCT GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTGTTAGC ATTTAGTGT TTGCGGACCT 3920
 CAGCTACAGC ATCCCTGCCA GGGCTGGGA CACCCAGAAG TTCTCGCCC CCTTACATCC AAACCTGAGC 3990
 AACCCGGAGT CTGGATTCTC GGGGACTCT CAGCTGTCTC GCGGGTTGTG CGGGGCCCGA GGCTGGAGGG 4060
 GGACCAAGTGG CGCTGTGGCT TCTACTCTG GCGCTGGAACT CGGGCTCTC AGCTCTGCCAG TCCGAGGGCTT 4130
 GGAGCCAGGT CGCTGGACCC CGAGGCTGCC CTCCACCTCTG TGCGGGGGGG ATGTCAGGAC ATGTTGGCT 4200
 CATCTGCCAG ACAGAGTGC CGGGGCCAGG GTCAAGGGCG TTGTGGCTGG TGTGAGGGCC CGGGTGGGG 4270
 GCCAGCAGGA GGGCCCTGGCT CCATTTCTCG ACGGGACCCGC CGGGGTGGGGT GATTAACAGA 4340
 TTTGGGTTGG TTGCTCATG GTGGGGACCC CTGCGCCCTC GAGAACCTGC AAAGAGAAAT GACGGGGCTG 4410
 TGTCAGGAG CCCAAGTGC GGGGAGTGT TGCAAGGGAGG CACTCCGGGA GTGCGCCGGT GCGCGTCCAG 4480
 GGAGCAATGC GTCTCTGGGT TGCTCCCCAG CGCGCTCTAC GCGCCTCCGT CCTCCCCCTC ACCTCCGGCA 4550
 TTCTGTGGTC CCGGAGCCCC AGCCCCCGC TCCGGACCTG GAGGGCAGCCC TGGGTCTCCG GATCAGGCCA 4620
 GCGGCCAAAG GGTGCGCCGCA CGCACCTGTT CCCAGGGCTT CCACATCATG GCGCTCCCT CGGGTTACCC 4690

5 / 15

Fig. 4 (Fortsetzung)

CACAGCCTAG GCCGATTCGA CCTCTCTCCG CTGGGGCCT CGCTGGCGTC CCTGCACCCCT GGGAGCGCGA 4760
GGGGCGCGCG GGCGGGGAAG CGCGGCCCAAG ACCCCCCGGT CCGCCCGGAG CAGCTGCCT GTCGGGGCCA 4830
GGCGGGGCTC CCAGTGGATT CGCGGGCACA GACGCCCAGG ACCGCCTCC CGACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCCTCG CGCGGACCCC GCGCGTCCC 4970
GACCCCTCCC GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCCC CCCCTCCCT TCCGCGGCCC 5040
CGCCCTCTCC TCGCGGGCGG AGTTTCAGGC AGCGCTGCCT CCTGCTGCC ACGTGGAAAG CCTGGGGCCC 5110
GGCACCCCCC GCGATG 5126

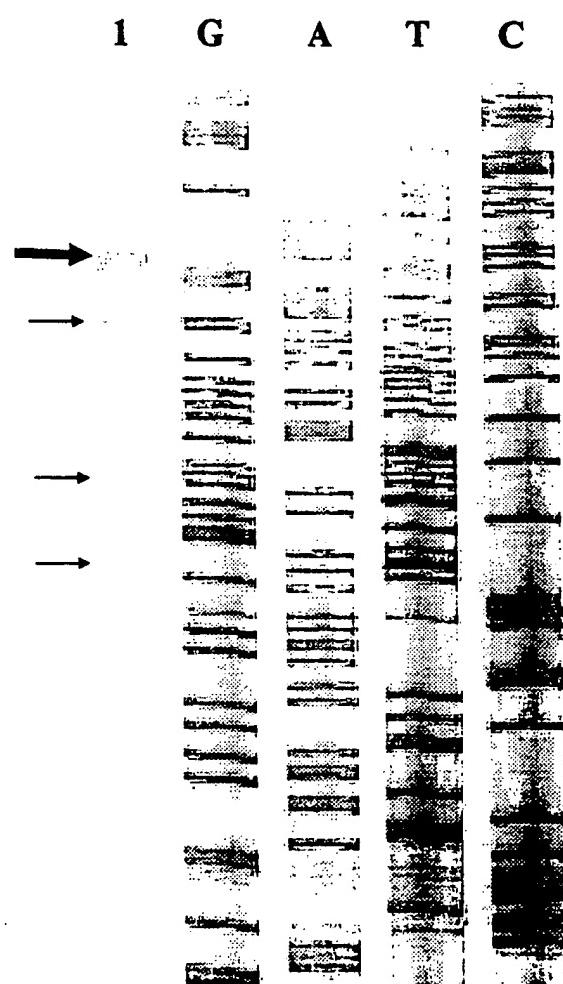
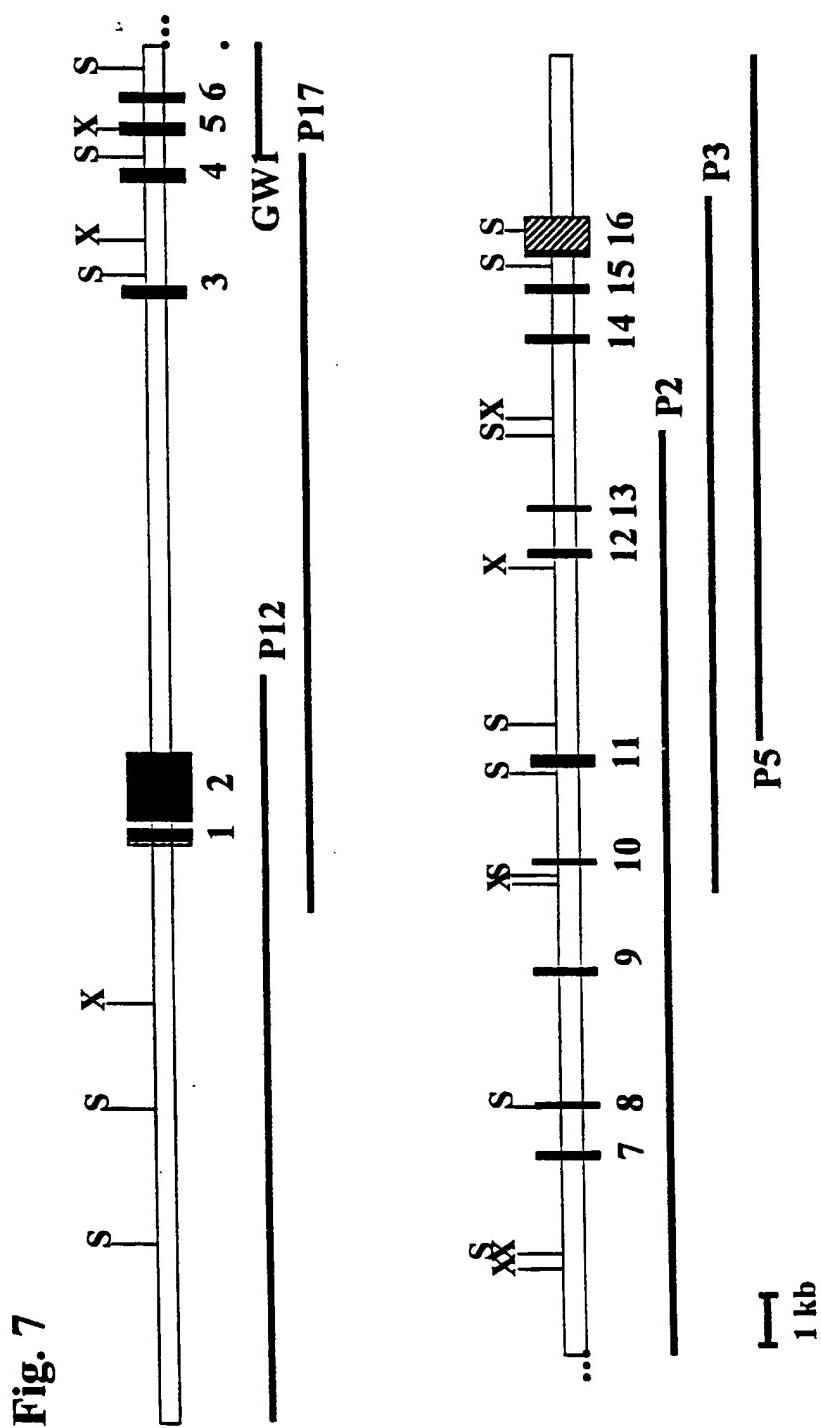
Fig. 5**ERSATZBLATT (REGEL 26)**

Fig. 6

GTTTCAGGCA GCGCTGGCTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCCG 70
 CGCTCCCGC TGCGGAGCCG TCGCTCCCT GCTGGCAGC CACTACCGC AGCTGCTGCC GCTGGCCAC 140
 TTCGTGGCGC GCCTGGGCC CCAGGGCTGG CGGCTGGTGC AGCGGGGAA CCCGGGGCT TTCCGGCGC 210
 TGGTGGCCA GTGCTGGGT TCGCTGCCCT GGGACGGCACG GCCGCCCCC GCCGGGGCTT CTTCCGCCA 280
 GGTGTCTGC CTGAAGGAGC TGGTGGCCCC AGTGTGCAAG AGGCTGTGCC AGCGGGGCCA AAAGAACGTG 350
 CTGGGCTTCG GTCTGGCGCT GTGGACGGG GCCCCGGGG GCCCCCCCAGA GGCCTTCACC ACCAGCGTGC 420
 GCAGCTACCT GCCCAACACG GTGACCGAGC CACTGGGGG GAGCGGGCGG TGGGGCTGCC TGCTGGCCCG 490
 CGTGGCGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC CGCTCTTTC TGCTGGTGGC TCCAGCTGC 560
 GCCTACCGG TGTCGGGCC CGCGCTGTAC CAGCTGGGC CTGCCACTCA GCGGGGGCC CGCCGACACG 630
 CTAGTGGACC CGGAAGCGT CTGGGATGCC AACGGGCTG GAACCATAGC GTCAAGGAGG CGGGGGTCCC 700
 CCTGGGCTG CGAGCCCCGG GTGCGAGGAG CGCGGGGGC AGTGCCACCC GAAGTCTGCC TTGCCCCAAG 770
 AGGCCAGGGC GTGGGGCTGC CGCTGAGCC GAGCGGGAGC CGCTGGGCCA GGGGGCTCTGG GCCCACCCGG 840
 GCAGGACGGC TGACCGAGT TCGCTGGTGT TCTGCTGTTG CTGACCTGCC AGACCCGGG AAAGAACCCAC 910
 CTCTTGGAG GTGCGCTCT CTGGCACCGG CCACCTCCAC CAATCGCTGG GCGGGGGCC CGACCGGGCC 980
 CCCCCATCCA CATCGGGCC ACCACGTCCC TGGGACACGC TTGTCCCCCC GGTGTACGCC GAGACCAAGC 1050
 ACTTCTCTA CTCCCTCAGGC GACAAGGAGC AGCTGGGGC CTCTTCTCA CTCACTCTC TGAGGCCCCAG 1120
 CCTGACTGGC GTCTGGAGGC TCTGGAGAC CATCTTCTG GTTCTCAGGC CCTGGATGCC AGGGACTCCC 1190
 CGCAGGTTGC CGCCGCTGCC CGACGCTAC TGGCAAAGTC CCACCTCTGG TCTGGAGCTG CTGGGNAACC 1260
 ACGGGCAGTG CCCCTACGGG GTGCTCTCA AGACGCACTG CGCGCTGGCA GCTGGGTCA CGCCAGCAGC 1330
 CGGTGTCTGT GCGGGGGAGA AGCCCCAGGG CTCTGTGGC GCCCCCGAGG AGGAGGACAC AGACCCCCGT 1400
 CGCCTGGTGC AGCTGCTCCG CGACACAGC AGCCCCCTGGC AGGTGTACGG CTTCGTCGGG GCCTGGCTGC 1470
 GCGGCTGGT GCGGGGGAGC CTCTGGGGCT CGAGGACCAA CGAACGGGGC TTCTCAGGA ACACCAAGAA 1540
 GTTCATCTCC CTGGGGAAAGC ATGCCAAGCT CTGCGTGCAG GAGCTGAGCT GGAAAGATGAG CGTGGGGAC 1610
 TGGCCTTGGC TGCGCAGGAG CGCCAGGGTTG GGCTGTGTC CGCGCCGAGA GCACCGCTGC CGTGAGGAGA 1680
 TCCTGGCAA GTTCTGCGAC TGGCTGATGA GTGTGTACGT CGTCGAGCTG CTCAGGTCTT TCTTTATGT 1750
 CACGGAGACC AGCTTCAAA AGAACAGGCT CTTTTCTAC CGGAAGAGTGT TCTGGAGCAA GTTCAAAGC 1820
 ATTGGAATCA GACAGCACTT GAAGAGGGTG CAGCTGGGG AGCTGCTGCC AGCAGGAGTC AGCAGGATTC 1890
 GGGAAAGCCAG CGCCGCCCCG CTGACGTCCA GACTCCCTT CATCCCCAAG CCTGACGGGC TGCGGCGGAT 1960
 TGTGAACATC GACTACGTG TGGGGAGCAG AACGTTCCG AGAGAAAAGA GGGCCGAGCG TCTCACCTGC 2030
 AGGGTGAAGG CACTGTTCA CGTCTCAAC TAGAGCGGG CGGGGGCCCG CGGCCCTCTG CGGCCCTCTG 2100
 TGCTGGCCCT GGACGATAATC CACAGGGCTT CGCGCACCTT CGTGTGCGT GTGGGGGCC AGGACCCGCC 2170
 GCCTGAGCTG TACTTTGCA AGGTGGATGT GACGGGGCGG TACGACACCA TCCCCAGGA CAGGCTCACG 2240
 GAGGTCTACG CGCACATCAT CAAACCCAG AACACGTACT CGCTGGCTGC GTATGGCTG GTCCAGNAGG 2310
 CGGGCCATGG GCACGTCCCG AAGGCTTCA AGAGCCACGT CTCTACCTTG AGACACCTCC AGCCGTACAT 2380
 GCGACAGTTC GTGGCTCACC TGCGAGAGAC CAGGGCGCTG AGGGATGCC CGTCATCGA CGRAGGCTCC 2450
 TCCCTGAATG AGGCCAGCG TGGCTCTTC GACGTCTTC TACGTTCTAC GTGCCACCCAC GCCGTGGCA 2520
 TCAGGGCCAA GTCTACCTC CAGTCCCAGG GGATCCCCAAGG GGGCTCATCG CTCTCACCGC TCTCTGCGAG 2590
 CCTGTGTAC GCGACATGG AGAACAAAGCT GTGGGGGGG ATTCGGGGG CAGGGCTGCT CTCGGGTTTG 2660
 GTGGATGATT TCTTGTGGT GTGACCTCAG CTCACCCACG CGAAAACCTT CCTCAAGGAC CGGGTCCGAG 2730
 GTGTCCCTGA GTATGGCTGC GTGGTGAAC TGCGGAAGAC AGTGGTGAAC TTCTCTGTAAG ACACGAGGC 2800
 CCTGGGTGGC ACAGGTTTTG TTCAGATGCC GCCCCACGG CTATCCCTT GGTGGGGCT GTGCTGGAT 2870
 ACCCGGACCC TGAGGTCGA GAGGGACTAC TCCAGCTATG CGGGACACTC CATCAGGCC AGTCTCACCT 2940
 TCAACCGGG CTTCAAGGCT GGAGGAACA TGCGTCGCAA ACTCTTGGG GTCTTGGGC TGAAGTGTCA 3010
 CAGGGCTTT CTGGATTTC AGGTGACAG CTCACCCAGG GTGTCGACCA ACATCTACAA GATCTCTCG 3080
 CTGCAAGCGT ACAGGTTTC CCGATGTGTG CTGCACTCC CATTTCATCA GCAAGTTGG AAAAACCCCA 3150
 CATTTCCTCCT GCGCGTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGCAGG 3220
 GATGTCGCTG GGGGCGAAGG GCGGGGGCCG CCTCTCTCCC TCCGAGGCC TGCGCACCAA AGTCTCACCT 3290
 GCATTCCGTG TCAAGCTGAC TGACACCGT GTCACTTCACTC GGGGTCACTC AGGACAGGCC 3360
 AGACCGAGCT GAGTCGGAAAG CTCCCCGGGA CGACGCTGAC TGCCACTGGAC CGCGCAGCCA ACCCGGCACT 3430
 GCCCTCAGAC TTCAAGACCA TCCCTGGACTG ATGGCCACCC GCCCACAGCC AGGCGGAGAG CAGACACCAG 3500
 CAGGGCTGTC ACAGCCGGGCT CTACGTCCTCA GGGAGGGAGG GGGGGCCAC ACCCAGGCC GCACCGCTGG 3570
 GAGTCGAGG CCTGAGTGA TGTTTGCCCG AGGGCTGCAT GTCCGGCTGAG AGGTGTAGTG TCCGGCTGAG 3640
 GCCTGACCGA GTGTCCAGCC AAGGGCTGAG TGTCAGGAC ACCTGCGCTC TTCACTTCCC CACAGGCTGG 3710
 CGCTCGGCTC CACCCCAAGG CGAGCTTTC CTCACTCCAGG GCGGGCTTTC CACTCCCCAC ATAGGAATAC 3780
 TCCATCCCCA GATTCGGCAT TGTTCACTCC TCGCCCTGCC CTCTTTGCC TTCACTCCCC ACCATCCAGG 3850
 TGGAGACCCCT GAGAAGGAGC CTGGGAATTG GAGTGACCAA AGGTGTGCC TGTACACAGG 3920
 CGAGGACCCCT GCAACCTGGAT GGGGGTCCCC GTGGGTCAAAT TTGGGGGGAG GTGCTGTGCC ASTAAAATAC 3990
 TGAATATATG AGTTTCTACG TTTTGAAGAA AAAAAGAAAAA AAAAAGAAAAA AA 4042

**Fig. 7**

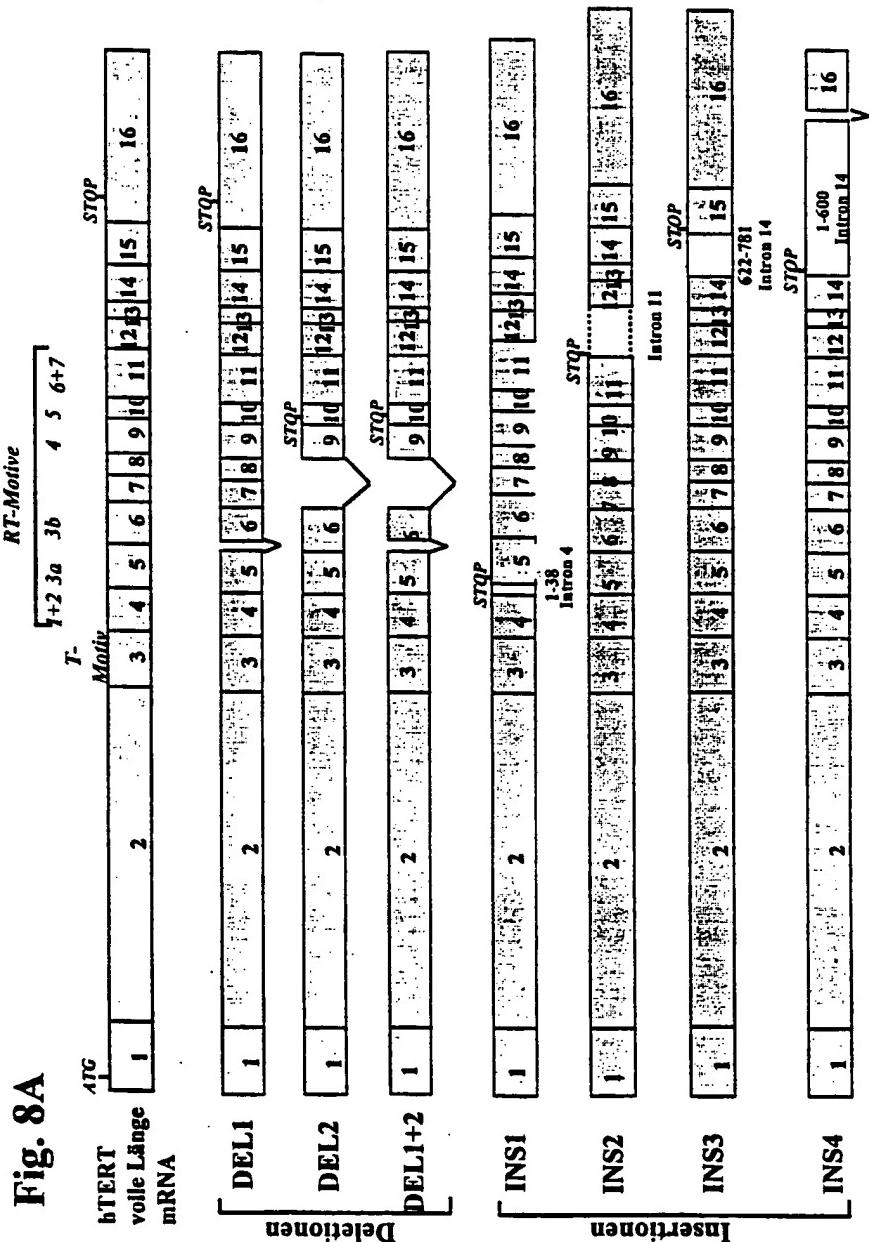
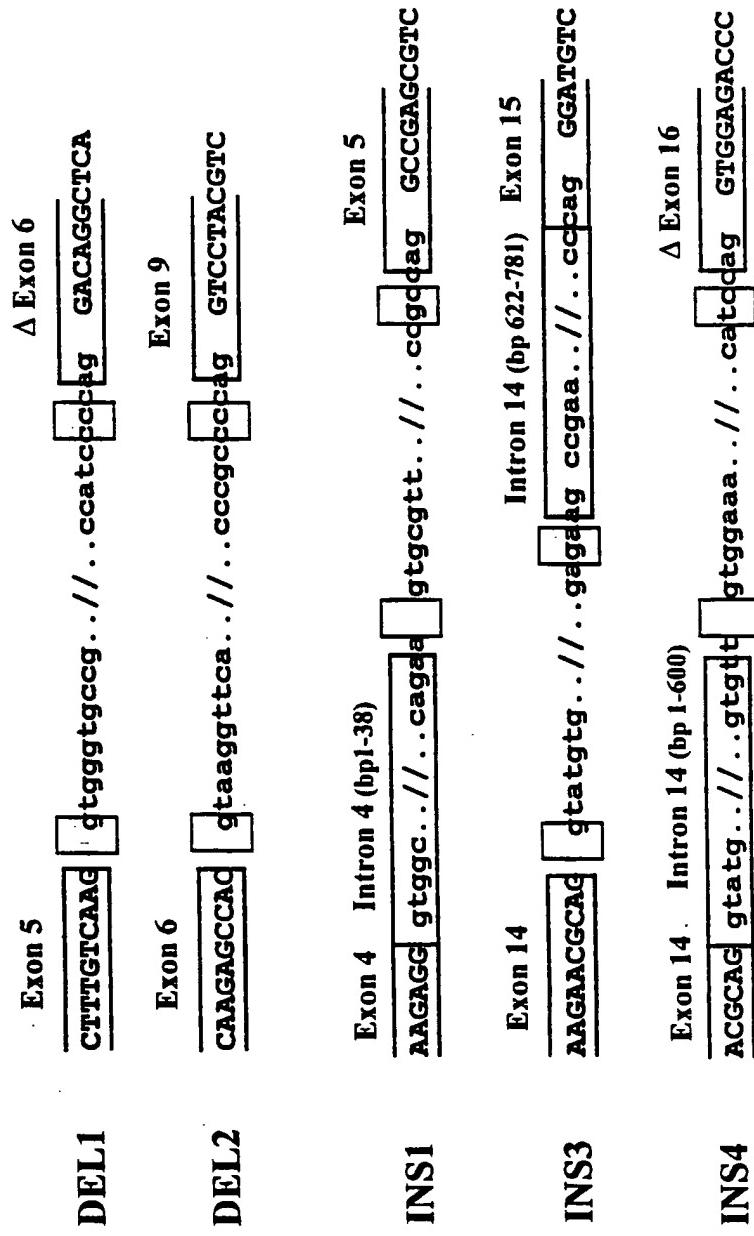


Fig. 8B

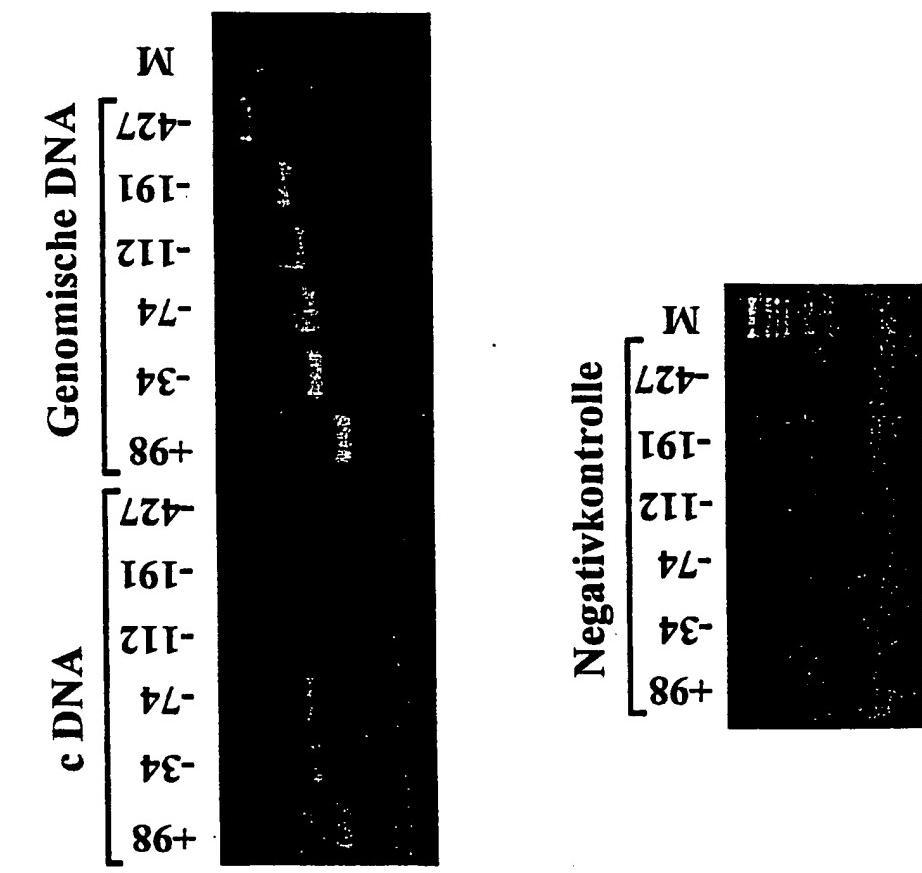


Fig. 9

Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCT GTCTAAAAAA AAAAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGG -11134
 ACAAAACAG AAATCAACAA CAAGAGGAAT TTGAAAACT ATACAACAC ATGAAAATTA AACATATAC -11064
 TTCTGAATGA CCACTGAGTC AATGAGAAA TAAAGAAAAGA AATTGAAAATTA TTATTAGG CAAATGATAA -10994
 CGGAAACATA ACCTCTCAA ACCCACCGTA TACAGCAAA GCAGTGTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GGCGCAGTGG CTCATGCCG TAACTCCAGC ACTTTGGGAG -10854
 GCCAAGGCGG GCAGATCGCC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTGTGCG -10784
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGCCACATG CTCGTAAATCC CAGCTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTGAAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGAT TGGCCCATG GACTCCAGCC -10644
 TGGGTAACAA GAGTGAACCC CTGTCCTCAA AAAAAAAAAA AAGTAGAAAATTA ATCTAAATAC ACAACCTAAT -10574
 GATGCACTT AAAGAACTAG AAAAGCAAGA GCAAACTAA CCTAATTTG GTAAAGAAA AGAAAATAATA -10504
 AAGATCAGAG CAGAAAATTA TGAAACTGAA AGATAACAAT ACAAAGATC AACAAAATTA AAAGTTGGTT -10434
 TTTTGAAGAAAG ATAACACAAA TTGACAAACCC TTTGCCAGA CTAAGAAAAA AGGAAAGAAG ACCTAAATAA -10364
 ATAAAGTCAG AGATGAAAGA AGAGACATTA CAACTGATCA CACAGAAATT CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTACACTA ATAAATTGAA AACATAGAA AAAATAGATA AATTCTTAGA TGCAACAC -10224
 CTACCAAGAT TGAAACCATGA AGAAAATCAA AGCCCCAACCA GACCAATAAC AACAAATGGGAA TAAAGCCAT -10154
 AATAAAAAGT CTCTAGCAA AGAGAAGGCC AGGACCAAT GGCTTCCCTG CTGGATTTA CCAATCATTT -10084
 AAAGAAGAAAT GAATTCACAT CCTACTCAAAT CTATTCGAA AAATAGAGGA AAGAATACTT CCAAACATCAT -10014
 TCTACATGGC CAGTATTACCT CTGATTCCAA ACCACAGACAA AAAACACATCA AAAACACAA AACAAAAAAA -9944
 CAGAAAGAAA GAAAACATCA GGCAACATAC CCTGTGAAT ACTGATACAA AAATCTCAA CAAAACACTA -9874
 GCAAAACAAA TTAAACACAA CCTTCGAAAG ATCATTCACTT GTGATCAAGT GGGATTATT CCAGGGATGG -9804
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAACACTA -9734
 TATGATTATT TCACCTTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCATGATA AAAACCTCA -9664
 AAAAACAGG TATAACAGAA ACATACAGGC CAGGACACATG GGCTCACACC TGCGATCCCA GCACTCTGG -9594
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTGAG ACTAGCTTG GCAACAAAT GAGACCTGGT -9524
 CTACAAAAAA CTTTTTAAAGA AATTAGGCCA GGCGATGATGG CATATGCCCTG TAGTCCCGC TAGTCTGGAG -9454
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCACTGAGC CATGAACATG TCACTGTACT -9384
 CCAGCCTAGA CAACAGAACAA AGACCCCCACT GAATAAGAAG AAGGAGAAG AGAAGGGAGA AGGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244
 AAGAAACATA TTCAACACATA ATAAAAGGCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAATGAA -9174
 AGCCTTCCCT CTAAGATCTG GAAAATGACA AGGGCCACT TTCACACTG TGATTCACAA TAGTACTAGA -9104
 AGTCTTAGCT AGAGCAATCA GATAAGAGAA AGAAAATAAA GGCACTACAA CTGGAAAGGA AGAAGTCAAA -9034
 TTATCCTGTT TGCAAGATGAT ATGATCTTAT ATCTGAAAAA GACTTAAGAC ACCACTAAA AACTATTAGA -8964
 GCTGAAATTG GGTACAGCAG GATACAAAAT CAATGTCACAA AAATCAGTAG TATTCTATA TTCCAACAGC -8894
 AAACATCTG AAAAAGAAC CAAAAGCA GCTACAAATA AAATTAACAA GCTAGGATT AACCAAAGAA -8824
 GTGAAAGATC TCTACATGAA AAATCTAAAAT ATGTTGATAA AGAAAATGAA AGAGGGCACA AAAAAGAAA -8754
 AGATATTCCA TGTTCATAGA TTGGAAGAAT AAATACTGTT AAAATGTCCA TACTACCAA AGCAATTTC -8684
 AAATTCAATG CAATCCCTAT TAAATACCA ATGACGTTCT TCACAGAAAT AGAAACACAA ATTCTAAGAT -8614
 TTGTACAGAA CCACAAAAGA CCCAGAACATG CCAAAAGCTAT CTCGACCAAA AAGAACAAAAA CTGGAAGCAT -8544
 CACATTACCT GACTTCACAT TATACACAA AGATGAGACA TGGACACAG GAAACAGATA TTTTCAACAC -8474
 AGATGAGACA TGGACACAG GAAACAGATA TTTTGACAA AGGTGCCAG AACAAATCC ATGCATCTAC AGTGAACATC -8404
 GGTGCAAG AACATACTTT GGGGAAAGA TAATCTCTTC AATAATGGT GCTGGAGGA -8334
 CTGGATATCC ATATGCAAA TAACAAATCT AGAACTCTGT CTCTCACCAT ATACAAAGC AAATCAAAAT -8264
 GGATGAAAGG CTAAATCTCA AAACCTCAA CTTGCAACT ACTAAAAGAA AACACCGGAG AACACTCTCA -8194
 GGACATTGGA GTGGGCAAGG ACTTTCTTGAG TAAATCCCTG CAGGCACAGG CAACCAAAGC AAAAACAGAC -8124
 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCCAAGCAA AGGAAACAAAT CAACAAAGAG AAGAGACAC -8054
 CCACAGAATG GGAGAATATA TTGCAAACAT ATTCACTCAA CAAGGAATTA ATAACCAAGTA TATATAAGGA -7984
 GCTCAACTA CTCTATAAGA AAAACACCTA CATTCTCAA AATAAGTCAT CAAATGGCA GAGAAATGCA ATAAAGCTGAT TTTCAAAAT AAGCAAAAGA TCTGGTAGA -7914
 AACACGGCT CAAACCTCAA CTTGCAACT ACTAAAAGAA AACACCGGAG AACACTCTCA -8194
 GAGAAATGCA AATAACAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAA ATGGCTTTA TTCAAAGAC -7774
 AGGCAAAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGAC ACTGTTGGTG GGAATGGAAA -7704
 TTGCTACAC TATGGAGAAC AGTTGAAAG TTCCCTCAA AACTAAAAT AAAGCTACCA TACAGCAATC -7634
 CCATTGCTAG GTATATACCT CAAAAAAGGG TCAACAAAGCT ATCTCCACTC CCACATTTC -7564
 TGCAGCACTG TTCACTAGCAG CCAAGGTTTG GAAGCACACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG -7494
 AAAATGTTG GTCACATACAA AATGGAGTAC TACGTCAGCA TAAAGGAGA TGAGATCTG TCAGTTGCAA -7424
 CAGCATGGGG GGCACCTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354
 CTCCCTTACT TGTGGGAGCA AAAATAAAAA CAATTGACAT AGAAAATAGAG GAGAATGGTG TTCTAGAGG -7284
 GGTGGGGGAC AGGGTACTA GAGTCAACAA TAATTTTATG TATGTTAA AATAACTAAA AGAGTATAAT -7214
 TGGGTTTTG TGAACACAAA GAAAGGATAA ATGCTTGAAG TGACAGATA CCCCATTTAC CCGATGTGA -7144
 TTATTACACA TTGATGCTT GTATCAAAT ATCTCATGTA TGCTAGATG AATAAACCTTA CTATATTAAA -7074
 AATTAAAATT TTAATGGCCA GGCACGGTGG CTCATGTCG TAACTCCAGC ACTTTGGGAG GCGGAGGC -7004
 GTGGATCACC TGAGGTCAAG AGTTGAAAC CAGTCTGCC ACCATGATGA AACCCCTGTCT CTACTAAAGA -6934
 TACAAAATT AGCCAGGCGT GTGGCACAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864
 TTGCTTGAAC CTGGGAGGCC GAGGTTGAG TGAGCCAGA TCATGCCACT GCACTGCAGC CTGGGTGACA -6794
 GAGCAAGACT CCATCTCAA ACAAACAA AAAAAGAG ATTAAAAATG TAATTTTAT GTACCGTATA -6724
 AATATATAC TCACTATATT AGAAGTTAA AATTAAAACA ATTAAAAG GTAATTAACC ACTTAATCTA -6654
 AAATAAGAAC AATGTATGTG GGGTTCTAG CTTCTGAAGA AGTAAAAGTT ATGGCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTAAC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGTTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGCA CCTCTGATCA -6374
 CACCGCTCTC TCATTACCG TGCTTTTTT CTTGTGTGCT TGGAGATTT CGATTGTTG TTCTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAATGGT GGTTGATTCC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTC ATCTTCAAGG GTCTCTGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGGCTGTG ACCGGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCTGCC -6094
 CGTTGGTGA CAGGGCATGA AGTGGCTTA TTACGCTT GCAAGAGTC CTCTGATPAC CATCTGGAAA -6024
 AGGGGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCTGC CAACCCAG GCCAGCAGT ATGGCCCCA -5954
 CCCGGGCGTGC TGCCAGAGGG AGAGGAGTC AGGCACCTCG AAGTATGGT TAAATCTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAAG CTTGAGTTAG GTGCCCTCTT TAAACAGAA AGTCATGGAA -5814
 GCACCCCTCT CAAGGGAAAA CCAGACGCC GCTCTGCCGT CATTACCTC TTTCCTCTT CCCTCTTCTG -5744
 CCCTCCGGT TTCTGATGCC GACAGAGTGA CCCCCGTTG GCTCTCTCG GCCCCGCTG AGGACCCCTC -5674
 TGCAAGGGC TCCACAGAC CGGGCCCTG AGAGGAGT CTGAGCTTG CTTATAACA AACTGGGATG -5604
 TGGCTGGGG CGGACAGCGA CGGGGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAAC CTTTCACAT -5534
 CCGAATGGAT TTGGATTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTCAAGG CTGAGAAAT -5464
 CCAAAGGGT AAACACAGGA CTGAGCTATG TTGCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTCCGCC CTAAGTACTT TTATGGTT TTACATAAGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGCCCTG CGGGCAGGGC ATGAGCACG ACAGCAGGAC CACTGACCGT CTCCTCTGG AGCTGGCAAC -5254
 ACAGCAGGAC CACTGACCGT CTCCTCTGG AGCTGGCAAC CGGGCAACAG CAGGCTGCGT -5184
 GTGACTCAGG ACCCCATACC GGCTTCTGG GCCCACCCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
 CCGTGGAACAC GAACATGACC CTTGCCCTGC TGCTTCCCTG GGTGGTCAA GGGTAATGAA GTGGTGTGCA -5044
 GGAATGCC ATGAAATTA CACGACTCTG CTGATGGGCA CGGCTCTTC CATCATATT CATCTTCACC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTC ATGAGCACG CACGGCTGCA CAAACATGAG TACAAACACC -4904
 ACTCTTTAC TAGGCCACA GAGCACGGSC CACGGCTCTG ATATATTAAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTCC TCTAGACTAG TAGACCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCCTGGT TGCTGCTTCC CGAGGGGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGTGCC ACAGTGGAC CAGGCCCTGC TCCACACCC CCGCTCTCG GCCTCAGCTT CTCCAGCAGC -4624
 TTCTTAAACCT CTGGGTGGGC CGTGTTCAG CGTACTGTC TCACCTGTC CACTGTGTC TGCTCTCAGC -4554
 ACGTAGCTG CACGGTTCTC CCTCACATGG GGTTGTGTC TCCTTCCCA ACACCATCAT GCGTTGAAGG -4484
 GAGGAGATTG TGCGCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGT CCGGGCCCC GATGCAGGTT -4414
 CCTGCGCTCC GGTGACACGC TGACCTCCAT TTCCAGGCGC TCCCGTCTC CTGTCATCTG CCGGGCCTG -4344
 CCGGTGTTT CTCTGTTT TGTTGCTCTG TCCACGTCA GCTGGCTGTG TCTCTGCCG CTAGGGCTC -4274
 GGGTTTTTA TAGCATAGG ACGGGGCGT GGTGGCCAG GGCGCTCTG GGAATGCAA CATTGGGTG -4204
 TGAAGTAGG AGTGGCTGTC CTCACCTAGG TCCACGGCA CAGGGCTGG GATGGGCCCG CCGCCAGGGA -4134
 CCCGCCCTC TCTGCCAGC ACTTCCCTGC CCCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATT GCCCCACAGC CCTGGAAATT -3994

c-Myc

CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCGG CTGTTTATT TTAATAGCTA -3924
CAAAAGCAGGG AAACTCCCTGC TAAATGTCC TTTAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854
GGACAGTTCC TCACAGTGA GAGGAACATG CGGTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
GACTCAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTC ACCCCATGGC -3714
AGGGGAGTGG TTAGGGGGT TAAGGGAGGT GGGGGCGG GCTGGGGCT ACTGACGCC CTTTTACTA -3644
AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCACTGTTG GGAGACTAAC CATAGGGAG TGGGGATGGG -3574
GGAAACCCGGA GGCTGTGCCA TCTTGTCCAT GCGGGAGTGT CCTGGGCAAGG ATAATGCTCT AGAGATGCC -3504
ACGTCTGTAT TCCCCAAAC CTGTGGACAG AACCCGCCG GCCCCAGGGC CTTGGCAGGT GTGATCTCG -3434
TGAGGACCTG GAGGTCTGGG ATCTTCTGGG ACTACCTGCA GGGCCGAAA GTAATCCAGG GTTCTGGGA -3364
AGAGGGGGG AGGAGGGTCA GAGGGGGCA GCCTCAGGAC GATGGACCA CTGAGCTGCA GGCTGAAAG -3294
GGAGGGAGGG CCTCGAGCCC AGGCTGCAGA GCGCCTCCAG AACCTGGAAA AAGCGGGGAA GGGACCCCTC -3224
ACGGAGCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGAC TCCGGCCTCC -3154
GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTTGGG ATTTGAGAA GCAACAGGAA -3084
ACCCATGCACT GTGAACTCA GGATTATTT AAAACAAAGG TTACAGAAA CATCCAAGGA CAGGGCTGAA -3014
GTGCCCTGG GCAAGGGCAG GGCAGGGCAG AGTGTATTG TTAGTATTG TTATTTTATT TACTTACTTT -2944
CTGAGACAGA GTTATGCTCT TTGTTGCCAG GCTGGAGTGC AGCGGCATGA TCTTGTCTA CTGCAACCTC -2874
CGTCCTCTGG GTTCAAGCAA TTCTCGTGC TCAGCTCTCC AAGTAGCTGG GATTTCAGGC GTGCACCAAC -2804
ACACCCGGCT AATTGTTAT TTTTAGTAGA GATGGGCTT CACCATGTT GTCAAGCTGA TCTAAATC -2734
CTGACCTCAG GTGATGCC CACCTCAGCC TCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
GGCTATTAA ACCATTAA AACCTCCCTC GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTT -2594
AATTCCCCCT TTACTCAGGA GTTACCCCTC TTTGATATT TCTGTAATT TCTGTGAGT GGGGATACAC -2524
CGTCCTCTGA CATATTACA GTTCTGTGA CCACCTGTTA TCCCAGGGG CCCACTGCAG GGGCAGCTGG -2454
GAGGCTGCAG GCTTCAGTC CCAGGGGGT TGCCATCTG CAGTAGAAC CTGAGTGTAGA ATCAGGGCGC -2384
AAGTGTGGAC ACTGTCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTAA AGTCCATCC -2314
CTCTACTCT ACTGGGATTG AGCCCCCTC CTATCCCCC CCAGGGGAG AGGACTCTC CTCACTCTG -2244
TGGAGGAAGG AATGATACTT TGTTTATTCTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTTGGTTTG -2174
TTTGTGTTGTTGTTGAGAGC GGTTTCACTC TTGTTGCTA GGCTGGAGGG AGTGAATGG CGCGATCTTG -2104
GCTTACTGCA GCCTCTGCC CCCAGGTTCA AGTGAATCTC CTGCTTCCG CTCCCATTT GCTGGGAGTA -2034
CAGGCACCCCG CCACCATGCC CAGCTAATTG TTGTTGTTT TAGTAGAGAC GGGGGTGGGT GGGGTTCAAC -1964

Fig. 10

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
 GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAATTAC TCTGTTAGA AACATCTGGG TCTGAGGTAG -1824
CAT-Box
 GAAGCTCACC CCACTCAGT GTTGTGGTGT TTTACCCA TGATAGAATT TTTTATTGT TGTTAGAAC -1754
 CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCG CTTTCAAAG ACACACTAAC TGCAACCCATA -1684
 ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCCG CCATGCACAT -1614
 GGTGTTAATT ACTCCAGCAT AATCTCTGC TTCCATTCTC TCTCTCCCT CTTTTAAAAT TGTGTTTCT -1544
 ATGTTGGCTT CTCTGCAGAG ACCAGTGTA AGCTACACT TAACTTTGT TGGAACAAAT TTCCAACC -1474
Spl
GCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG -1404
 GGATTCTAG AAGAGCGACC TGTAATCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334
 GCCCAGGGAG CGTGCAGGGC CTGTTCAAAT CCTAGCTCCA TAAATAAAGC AATTTCCCTC GGCAGTTCT -1264
 GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTAGC ATTCAGTGT TTGCCGACCT CAGCTACAGC -1194
 ATCCCTGCAA GCCCTCGGGA GACCCAGAAG TTTCTGCC CTTAGATCC AAACTTGAGC AACCCGGAGT -1124
 CTGGATTCTT GGGAAAGTCCT CAGCTGTCC CCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACCAAGTGG -1054
 CCGTGTGGCT TCTACTGCTG GGCTGGAAAGT CGGGCCTCCT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984
 GCCTGGACCC CGAGGCTGCC CTCCACCCCTG TCCGGGGGGG ATGTGACAG ATGTTGGCT CATCTGCCAG -914
 ACAGAGTGGC GGGGCCAGG GTCAAGGCG TTGTGGCTGG TGTGAGGCGC CGGGTGCAG GCCACAGGA -844
CCAC-Box
Spl
 GCGCCTGGCT CCATTCCCA CCCTTCTCG ACGGGACCC CCCGGTGGGT GATTAAACAGA TTTGGGGTGG -774
 TTTGCTCATG GTGGGGACCC CTCGCGCCCT GAGAACCTGC AAAGAGAAAT GACGGGCCTG TGTCAAGGAG -704
 CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCTG GCCCCTCCAG GGAGCAATGC -634
AP-2
 GTCCCTGGGT TCGCCCCAG CCGCGCTAC GGGCCTCCGT CCTCCCGT ACGTCCGGCA TTGGTGGTGC -564
 CGGGAGCCCG ACCCCCCCGCG TCCGGACCTG GAGGCAAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494
 GGTCGCCGCA CGCACCTGTT CCCAGGGCTT CCACATCATG GCCCCTCCCT CGGGTTACCC CACAGCTAG -424
Spl
 GCGCATTGCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCCT GGGAGCGCGA GCGGCGCG -354
Spl
GGCGGGGAAG CGGGCCCGAG ACCCCCCGGGT CCGCGGGAG CAGCTGCCTG GTCGGGGCCA GGCCGGGCTC -284
c-Myc
 CCAGTGGATT CGGGGGCACA GACGCCAGG ACCGGCTCC CCACGTGCG GAGGGACTGG GGACCCGGGC -214
Spl
 ACCCGTCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCC GCCCGTCCC GACCCCTCCC -144
Spl
 GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCTT CCCGTCCCT TCCGCGGCC CGCCCTCTCC -74
c-Myc
TCGCGGCCG AGTTTCAGGC AGCGCTGCCT CCTGCTGCC ACGTGGAAG CCCTGGCCCC GGCCACCCCC -4
 GCGATG

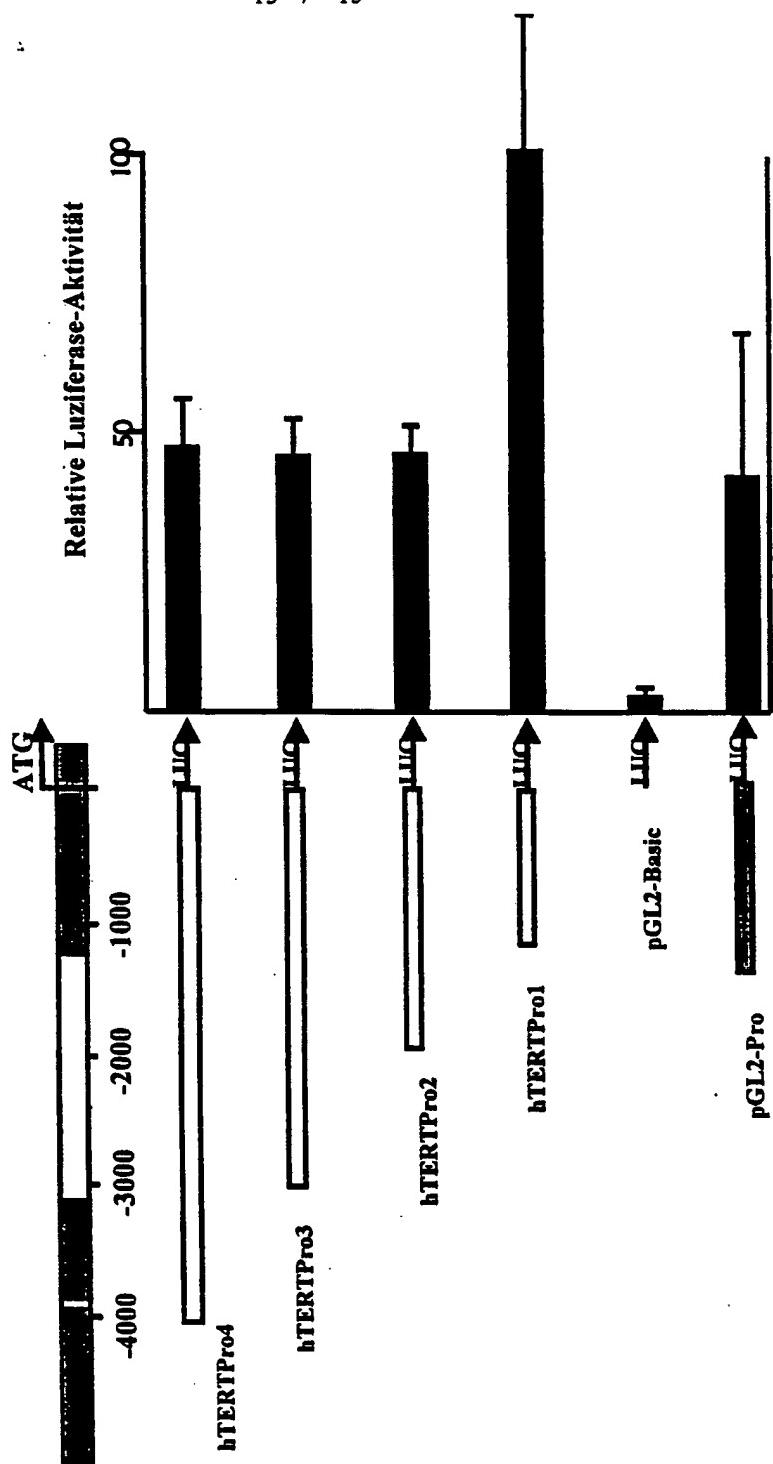


Fig.: 11

ERSATZBLATT (REGEL 26)

SEQUENZPROTOKOLL

	atacaaacac	atgaaaattt	aacaatac	ttctgaat	ccagtggat	aatgaagaa	240
5	ttaaaaagga	aattggaaaa	tttatttaag	caaataata	cggaaacata	acctctaaa	300
	acccacggta	tacagcaaaa	gcagtgtta	gaaggaaat	tatagctata	agcagctaca	360
	tcaaaaatgt	agaaaaggca	ggggcgatgg	cttcatttt	taatcccag	acttgggag	420
	gccaaggcgg	gcatatcgcc	tgaggctcagg	agttcgagac	caggcttacc	aacacagaga	480
	aaccttgtcg	ctactaaaaa	tacaaaattt	gtggggcatg	gtggccatcg	cctgtatatt	540
10	cagctactcg	ggaggcgttag	gcaggataac	cgcttgaacc	caggaggctgg	agtttgcgg	600
	gagccggat	tgcgccattt	gactccagcc	tgggtacaaa	gagtggaaacc	ctgtctcaag	660
	aaaaaaaaaa	aatgtaaaaa	actttttaaaat	acaacctaata	gtgcacccct	aaagaacttag	720
	aaaagcaaga	gcaaaactaa	cctaaaattt	gtaaaaaaa	agaataataa	aatgtccagg	780
	cagaaaaaaa	tgaaaatggaa	agataacaaat	aaaaaaagatc	aaaaaaatta	aaatgtggg	840
15	ttttgaaaag	ataaacaaa	tttgcacaaa	tttgcacca	ctaaagaaaa	agggaaaaaa	900
	acctaataaa	ataaaagtca	agatggaaaa	agagacatta	caactgtata	cacagaaattt	960
	caaaggatca	ctagggtca	ctatgagca	ctgtacacta	ataaatttga	aaaccttagaa	1020
	aaaatagata	aatttcttga	tgatcatcaac	cttaccaat	tgaaatcgat	agaadatccaa	1080
	agcccaaaaca	gaccaataac	aataatggga	ttttaaggcat	ataaaaaatgt	ctcttagca	1140
20	agagaagccc	aggacccat	ggcttccctgg	ctggatattt	ccaaatctttt	aaagaaatgt	1200
	gaatttcaat	cctactcaaa	cttattctga	aaatagagga	aagaataactt	ccaaactcat	1260
	tctacatggc	cagtattacc	ctgttccaa	aaccagacaa	aaacacatca	aaaacaaaca	1320
	aaaaaaaaaa	caaaaaagaa	gaaaactaca	ggccaaatcc	cttgcgtatgg	actgtatca	1380
	aaatccctca	caaaacacta	gcaaaaaacc	tttaaacaaca	ccttcggaaag	atcatttatt	1440
25	gtgtatcaatgt	ggggattttt	ccagggtatgg	aggatgttt	caacatgtc	aaatcaatca	1500
	atgtgtatca	tcatcccaac	aaaatgtatgt	acaaaaacta	tatgtatttt	tcactttatg	1560
	cagaaaaagc	atttgataaa	attctgcacc	tttgcata	aaaacccctca	aaaaaccagg	1620
	tatacaagaa	acatacaggc	caggcacat	ggcttcacacc	tgccatcccc	gcactctggg	1680
	aggcccaagggt	ggggatgttt	cttggggccca	ggggatgttag	actaggcttgg	gcaaaaaat	1740
30	gagacccgtt	ctacaaaaaa	ttttttttaa	aaattagccca	ggcgtatgtt	catacgctt	1800
	tagtccccagc	tagtctggag	gtcgagggtgg	gagaatctact	taactcttgg	agggtcgaggc	1860
	tgcagtgtgc	catgaacat	tcactgtact	ccagccat	caacagaaca	agacccccact	1920
	gaataagaag	aggagaagg	agaaggggaga	aggggagggg	aaggggaggag	gaggagaagg	1980
	aggagggtgt	ggagaatgtt	ggggggaaagg	ggggggaaagg	ggggggaaagg	aagaacacata	2040
35	tttcaacata	aaaaaaagccc	tatatgcac	accggatgtt	tattatgtt	aaaaaaactgtt	2100
	agcccttccct	ctaaagatct	gaaaatgtca	aggggccactt	tttccacact	tgatccaa	2160
	tagtactaga	agtccctact	agagacatca	gataagagaa	agaataaaaa	ggcatccaaa	2220
	ctggaaagga	agaagtcaaa	tttccctgtt	tgcatgtat	atgtatttat	atctggaaaa	2280
40	gacttaaagc	accactaaaa	aacttattga	gtcgaaattt	ggtacagcag	gataaaaaat	2340
	caatgttacaa	aaatctgtat	tatttctata	tttccaaac	aaacaatctg	aaaaaaagac	2400
	caaaaaaaagca	gctcaaaaat	aaataaaaaa	gcttagaaat	acccaatggaa	gtggaaatgc	2460
	tctacaatgt	aaacttataaa	atgttgcata	aagaatttga	agggggcaca	aaaaaaagaaa	2520
45	agatatttca	tgttcataga	tttgcggaaat	aaataactgtt	aaaatgtcca	tactacccaa	2580
	agcaattttc	aaattcaatg	caatccctat	aaaaatacta	atgcattttt	tcacagaaat	2640
	agaagaaaaac	atttcaatgt	ttgtacagaa	ccacaaaaaa	cccacaaat	ccaaagatct	2700
	cctgtacccaa	aaaagaaaaaa	tttgcggatcc	caatccatct	gacttccat	tataactcaa	2760
	agcttataatgt	acccaaacta	catgttact	gcataaaaaac	agatgagaca	ttggaccagag	2820
50	gaacagaata	gagaatccag	aaacaaatcc	atgcatttac	agtgtactca	tttttgacaa	2880
	agggtccaaag	aacatacttt	ggggaaaaga	taatctttt	aataaatgtt	gttggggagaa	2940
	cttgatattcc	atatgtcaaa	taataactact	agaatctgt	cttcattttat	atacaaaagc	3000
	aaatcaaaaatgt	ggatggaaa	ctttaatcttca	aaacctttttt	ctttgcact	actaaaaaaa	3060
	aaacacggag	aaacttctca	ggacatttgg	tttgcggaaaag	tttttgcact	tttttttttt	3120
55	caggcacagg	caacaaagc	aaaaacagac	aaatgggatc	atatcaatgtt	aaaaagcttc	3180
	tgcccgacaa	aggaaacaat	caacaaagag	aagagacaaac	ccacacat	ggagaatata	3240
	tttgcacaaat	atttcatcttca	caagggat	ataacatgtt	ttatataagg	gtccaaatca	3300
	cttctataaga	aaaacaccta	ataagcttta	tttcaaaaaat	aagcaaaaat	tctgggttag	3360
	catttctca	ataatgtcat	acaaatggca	aacaggcatc	tgaaaaatgt	ctcaacacca	3420
60	ctgtatcatca	gagaaatgtca	aatcaaaaaact	actatgttgc	atcatcttcat	cccaatgttca	3480
	atggctttta	ttcaaaagac	aggcaataac	aatgtccat	gaggatgtgg	ataaaaaaggaa	3540
	accccttggac	actgttgggt	ggaatggaaa	tttgcatttt	tatgttgcac	agtttgcgg	3600
	ttccttcaaa	aactaaaaaa	aaatctgttca	tacacatc	ccatgttgc	gttatatactc	3660
	aaaaaaagg	aatctgttca	tcaaaatgttca	atctccat	ccacatcttac	tgccatgttgc	3720
65	tttcatatgt	ccaagggtttt	gaagcaactt	cagtgttccat	caacacacg	atggggaaa	3780
	aaaatgtgtt	gcacatcac	aatggatgtt	tacgcacggca	taaaaaaaa	tgagatcttgc	3840
	tcagtgtca	cagcatgggg	ggcactgttgc	agttatgttt	gttggaaata	ccaggcacag	3900
	aaagacaaaac	ttttcatgtt	cttccatcttact	tgttggggac	aaaataaaaaa	caatgttgcac	3960
	aaaaatgtatgt	gagaatgggtt	tttgcgtatgg	ggggggggac	agggttgcata	gagtcacaaa	4020
	taattttatgt	tatgttttaa	aaaaactttaaa	agatgtataat	tttttttttt	gtacacacaaa	4080

5 gaaaggataa atgcttgaag gtacagata ccccatatt cctgtatgtga ttattacaca 4140
ttgtatgcct gtatcaaat atctcatgtc tgctatagat ataacccta ctatattaaa 4200
aattaaaatt ttaatggcca ggcacgggtt ctcatgtccg taatcccage actttgggag 4260
gcggaggcgg gtggatcct tgaggcttgg agtttggaaa cagtctggcc accatgtatg 4320
aaccctgtct ctactaaaga tacaaaaat aggccggcgtt gggtgcacat accttgatg 4380
5 ccaactactc aggaggctgtt gacggaggaa tgcttgcac tcggaggcgc gaggttgcag 4440
tgagccgaga tcatggccat gcactcgacg ctgggtgaca gaccaactt ccatctcaaa 4500
acaaaaacaa aaaaaagaag attaaaattt tgatccatgata aatataact 4560
ctactatatt agaaggtaaa aaaaaaaca attataaaag gtaataaacc acttaatct 4620
10 aaataaagaac aatgtatgtt gggtttctag ctctgttgaaga agttaaaatg atggccacga 4680
tggcggaaat gtgaggaggag aacagtggaa tgacttgtt tgtagcgttc atactctct 4740
taatgtactt aatttttaacc aaaaaaaaaaaggc tggggagaatg ttaaaggcga ttctataa 4800
cctaaaaacaa ctgctaataa tggtgaaagg taatcttat taattaccaa taattacaga 4860
tatctctaaa atcgagctgc agaattggca cgctgtatca caccgtccct tcattcacgg 4920
15 tgctttttt ctgtgtgtt tggagattt cgattgtgtt ttgcgttggg ttaaactta 4980
atctgtatga atccctgttga aaaaaatgtt tggttattcc tccagaagaa tttagtacc 5040
tggcggaaag cagggtggctc tggtggactt acggacttca atcttcagg gtcttggcc 5100
aagacccagg tgaaggcag aggccgtatg accccggggc agggaaagtc ggatgggaaag 5160
gggcgtatgag aaggctgtt cgttgggtt cagcgtatga agtgcctta ttacgtttt 5220
20 gcaaaagattt ctctggata catctggaaa agggggccag cgggaatgtca aggagtca 5280
agecttcttc tcaaaccctc gccagcaget atggccccc cccggcggtg tgccagggg 5340
agagggtca aggcacccctc aagtatgtt tttacatgtt ttccatctgtt agcgttgacc 5400
aagggttatt ctgaggaaat ctgtgtttag tggtttttt taaaacagaa agtcatggaa 5460
gcacccctct caaggaaaaa ccagaccccc gctctcggtt cattacetc tttctctct 5520
25 ccctctcttg ccctcgccgt ttctgtatcg gacagatgtt ccccccgttga gcttcccg 5580
gccccgtgtc aggacccctt tggcaaaaggc tccacagacc cccggccctgg agagaggagt 5640
ctgagccctgg ctaataaca aactgggtt tggttggggg cggacagcga cggccgggatt 5700
caaaagactt atccccatgat taaaatccat cttttccat ccaatgtt tggtttttt 5760
30 tcttaatatt ttcttaaattt tcatcaataa acattcagga ctgcagaaat ccaaaggcgt 5820
aaaacaggaa ctgagctatg tttgccaagg tccaaggact taataaccat gttcagagg 5880
atttttccgc ctaacttactt ttattgtt ttctataatg ggtttaggtt gcaaggaaaa 5940
gtacacggag agaggcttgg cggcagggtc tatgacccacg cggggccac cggggagaga 6000
gtccccggcc tggggggctt acagcaggacg cactgttccctt cttcccttggg agtcccccaca 6060
ttgggcaacg cgaaggccgc cacgtgtgtt gtgtacttggg accccatacc ggcttccctgg 6120
35 gccccccac actaaccctcga gaagtccatgg agctcttgc acctggaaac gAACATGACC 6180
cttgcctgc tggcccttgg ggtgggtcaaa gggtaatggat gttgggtgtca gggaaatggcc 6240
atgttaaaaat caccactctg ctgtatgggg cccgttccctt catcttattt catttcacc 6300
cccaaggactt gaatgttcc acgaacttctt cttgggtgtt gcaaggatgtt caaaacttgc 6360
tacaaacacc actcttttac tagggccaca gaggcaggsc cacacccctg atatataag 6420
40 agtccaggag agatgaggatg gttttcagcc accaggctgg ggtgacaaca gcccgtgt 6480
agtcgttcc tctagacttag tagaccctgg cggcacttcc cccagattt aggccctgtt 6540
tgctgttcc cggggccgc acatggccctg gagactcgtc ctgggtgtcc acacttgaggc 6600
cageccctgc tccacaccctt cccgttccatg gcctcgttcc cttccacggc ttcccttacc 6660
ctgggttggcc cgtgttccatg cgttactgttcc tccatgttcc cacttgttct tggccatcg 6720
45 acgttagctcg cacgggttcc tccatcatgg ggtgttgc tccctccca acactcacat 6780
gcgttgaagg gaggagatc tgccgttccccc agactggcgc ctctgtggcc gaaacctggc 6840
cgtggcccccctt gatcgagggtt cctggccgtc ggctgcacgc tgacttccat tccaggccgc 6900
tcccccttc tctgtatctg cggggccctg cgggtgtttt ctgttcttctt tggccatcg 6960
tccacgttca gctgtgtgtt tctgtccccc ctgggttcc ggggtttttt taggtatagg 7020
50 acggggccgt ggtggggccag ggccgttcc gggaaatgtca cattttgggtt tgaaaggtagg 7080
agtgcctgtc ttcacacttgg tccacggca caggccctgg gatggagccc cccggccagg 7140
cccgcccttc tctgtcccaacg acctttctgc cccctccctt ctggaaacaca gagttggcagt 7200
ttccacaaacg actaaggatc ctetttccaa aagaccacgc attggccacc cttggatcc 7260
gccccccacgc cctggggatc cacgtgtacta cgcacatcat gtacacactc cccgttccacga 7320
55 ccgaccccccctt ctgttttttatt ttaatagctt caaaaggaggaa aatccctgc taaaatgtcc 7380
tttaacaaac tgggttaaaca aacgggttca tccgcacggt ggacatgttcc tcacagtgtt 7440
gaggaaatctg cctgtttataa agcctgttgg cttccatggaaatctgtt gatggaaacaa 7500
tgcaccccttc atgggataacg tgcgtatgtt gtcataaaag aaaaatccatccatggcc 7560
agggggatgg ttaggggggtt taaggacggt gggggccgc gctgggggtt actggccacggc 7620
60 ctttttacta aagccatgtt cctgggttcc atggatattgg ctcaatgtatg ggagactaac 7680
catagggggatgg tggggatgg ggaacccgggaa ggctgtgc tcccttgcattt gcccggatgt 7740
cctggccaggataatgttcc agagatgtcc acgttccatg tcccccaacat tggccatcg 7800
aaccggccccc gccccccaggc ctttgcagggtt gtgtatctgg tgaggacccctt gagggtctggg 7860
atccccctggg actatctgtca gggccggaaa gtaatccagg ggttttggggg agaggccgggg 7920
65 aggagggtca gagggggggca gcttcaggac gatggaggca gtcagttca ggttttggggg 7980

6 / 18

gtcggatgg tgcaggctcg gggtgaggtc gccaggccctt tggtgagctg gatgtgcggg 7500
 gtctgcatgg tgcaggctcg gggtgaggtc gccaggccctt tggtggctg gatgtgtgg 7560
 gtcggatgg tgcaggctcg gggtgaggtc gccaggccctt tggtgagctg gatgtgcggg 7620
 5 tgctggatg tgcaggctcg ggggtggatg agccaaggcc ttccggatgc tggatgtggg 7680
 gtgtccggat ggtgcaggctcg ggggtggatg cttccggatgc tggatgtgg 7740
 ggtgtccggat tggtgcaggctcg cccgggtggatg gtacccaggcc cctccggatgc tggtggatgtg 7800
 cggtgtctgg atgggtgcagg tccgggtggatg ggtgcggccagg ccctgtctgt agctggatgt 7860
 gctgtatccg gatgggtgcag gtcgggggtg aggtgcggccag gcctcgcaatg gagctggatg 7920
 tgctgtatcc ggtatggtgcg ggtctggatg aggtgcggccag gecctgcgg ttagtggatg 7980
 10 atgcgggtgcg ggtatggtgcg ggtctggatg aggtgcggccag gecctgcgg ttagtggatg 8040
 gtgcgggtgtc cggatggatgc aggtctgggg tgagggtgcggcc aggcctgtgt gtgagctgg 8100
 tggtgtctat cccgatggatg cagggtccggg tgagggtgcggcc aggcctgcgg tggtggatgtg 8160
 atgtgtgtat tccggatggatg gcaaggatcg cgtgaggatcg ccaggccctg cgggtggatgtg 8220
 15 gatgtgcagt gtacggatgg tgcaggctcg ggggtggatg gccaggccctg gecgtggatgt 8280
 gatgtgtgt gttctggatg gtgcaggctcg ggggtggatg cccaggccctg tggtggatgt 8340
 tggatgtgtg tggatctggatg gtcggatgc cgggggtggatg tccggatgc ctcggatgc 8400
 tggatgtatcg tggatcccgt gtcggatgtt gtcggatgttca ggggtggatgc gccaggccctg 8460
 tggatggatgt gatgtgcggatg tggatgtgttca ggggtggatgc gccaggccctg 8520
 20 tggatgtgtg gatgtgcggatg tggatgtgttca ggggtggatgc accaggccctg 8580
 cgggtgatcg gatgtggatg tgccttcgtc ttaag 8616

<210> 6
 <211> 2089
 <212> DNA
 25 <213> Homo sapiens

<400> 6
 gtaactgtatc cccacgcccag gcctctgttt ctcaagaatcc tggaaacacca gccccggctc 60
 30 agcatgcgcc tgcctccact tgcctgtgt tccctggctg tgcagctctg ggctgggagc 120
 cagggggcccc gtcacaggcc tggtcccaatg ggattctgtg caaggctctg actgccttgg 180
 gtcacacgttc tcttacttgtt aaaaatcgaa gtttgcacaatgg agtggctctt aggggttgtt 240
 aacgcagaagg gatTTaaatc agatggaaac actaccacta gcctcccttg cttttccctgg 300
 gatgtggatgc tggatcttc tttttttttttt tgatgtggatgc tctactctg 360
 ttggccggcc tggaggatcg tggataataatgg ttcggatctact gcaacttcca cctctgggtt 420
 35 ttaaggcatt caccaggctt agcctcttaa gtagctggga ttacaggcac ctggccaccac 480
 gcctggctaa tttttgtact ttttaggagag acgggggttcc accatgttgg ccaggctgtt 540
 ctccgaactca tgacccatgg tggatccatccc accttggctt cccaaatgtgc tgggtttaca 600
 ggctaaaggcc cctgtggccat ccccccattt tttttttttt catgtgttgc tttttttttt 660
 ttcaatctat tggatgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720
 40 cagggagcac ctgtgcagggg agcacatggg gataggagag ttcccccattt agttaacttcc 780
 taggtggctg catttgaatg gctgtggatg tttgtctgtca atgtttggctt gatgtggatgt 840
 ttaggtgttgc acagattcaa gctggatgtt catcgtgtt gacccatggccatccatgg 900
 gatgtggccat cctgtgttgc gggatgttgc gatgtggatgtt gatgtggatgtt gatgtggatgtt 960
 45 tgactgtggaa ggggttttagt cagaatggatc ggggttttttcc agtcccttgc caactctgg 1020
 tccctgggggg gccttgcacccatggcc ccaaatcagggtt atgttgcacccatggcc 1080
 cagcagacatc cgtcaggatgtt aacacaggccctt ctgggtctggg gaccccgacccatggcc 1140
 gcccatttccat tgcattctggg ggagggttgc gggatgttgc gatgtggatgtt gatgtggatgtt 1200
 aatgcacccatc acttagactt tacacgttatt taatgtgttgc cggccatggatgtt gatgtggatgtt 1260
 50 accaggatgtt tggaaatggatc ttaatgtgggg tgacccatggccatggccatggccatggcc 1320
 ccccaagatgtt ctccctgtca ctactggggatc ttttttttttctgttgc ctttttttttctgttgc 1380
 cccctctccctt ggacagggttca cctgtggccatggccatggccatggccatggccatggccatggcc 1440
 agggcaccatgg ctccggatggccatggccatggccatggccatggccatggccatggccatggccatggcc 1500
 acagatgtccatc aggttccatggccatggccatggccatggccatggccatggccatggccatggccatggcc 1560
 55 gggaaaaaggcc caaggccatggccatggccatggccatggccatggccatggccatggccatggccatggcc 1620
 tccctgggttgc agtcccttgc ttttttttttctgttgc ctttttttttctgttgc 1680
 ctttttttttctgttgc ttttttttttctgttgc ctttttttttctgttgc 1740
 gggacaggatgtt ccctgtgtggggggatggccatggccatggccatggccatggccatggccatggccatggcc 1800
 caggctccctt ggtgtgtatgtt gttggatgttgc gggatgttgc gggatgttgc gggatgttgc 1860
 tccccaggatgtt tgactatagg accaggatgttgc ctttttttttctgttgc 1920
 60 agggcaccatgg ctccggatggccatggccatggccatggccatggccatggccatggccatggccatggcc 1980
 tgggtggccctt gggccatggccatggccatggccatggccatggccatggccatggccatggccatggcc 2040
 tagtctgttgc ttttttttttctgttgc 2089

```

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens
5
<400> 7
gtggctgtgc tttggtttaa cttcctttt aaacagaagt gcgttgagc cccacatgg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag gccatggca 120
cgccgcacaaac ccatttgtgc gcacagttag gtggccgagg tgccgtgccc tccagaaaaag 180
10 cagcgtgggg gtgttagggg agctcttggg gcaggacag gctctggagg ccacaagaag 240
cagcggggc aggggcttgg tgcagcacgg cccgaggtcc tggatccgtg tcctgtgtg 300
gtgcgcagcc tccgtgcgttacg gggccccggg accaggccac gactggcagg 360
agcccaacccgg gctctgaggta ttctggaccc tggccacgg ctccgtcacc ccacccctgt 420
ggctgggggt gtcgtggta ccccgatc tggaggaggt gtgggggtgag gtggacagag 480
15 gtgtggcatg aggatccccgt gtgcacacaca catgcggcca ggaacccgtt tcaaacagg 540
tctggggaaat ctgggggggg ttcttagtcc cgggtctggg tggctgggaa cactggggag 600
gggctgttcc tccccctgggt ccctatggg ggggtgggac ttggccggat ccactttctt 660
gactgttcc catgtgttcc cgccag 687
20
<210> 8
<211> 494
<212> DNA
<213> Homo sapiens
25
<400> 8
gtgggtgccg gggacccccc tgaggcagccc tgctggaccc tgggagtggc tgccgtattg 60
gcacccatcg ttgggtggag gaggtaatcc tgggtggcc gcaggaggatg caggtagccc 120
tgtcaactgtt gaggacacac ctggcaccta gggtgaggc cttcagccctt tcctgcagca 180
catggggccg actgtgcacc ctgactggcc gggctcttat tcccaaggag ggtcccaactg 240
30 gattccagt tccgtcagag aaggaacccgc aacggcttag ccaccaggcc ccgggtccctt 300
gcaccccaagt cctgagccag gggtctcttgc ttctggggct cagagagggg acacagcccg 360
ccctggccctt ggggtctgg tgggtggggg tcaagagaggt agtgggggac accggcaggc 420
caggccctga gggcagaggt gatgtcttag ttctgtgtg gcaactgtca gtctcttcgc 480
40
35
<210> 9
<211> 865
<212> DNA
<213> Homo sapiens
40
<400> 9
gtaaagggtca cgtgtgatag tgggtgtccag gatgtgtgtc tctggatata gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttctt gtgggtggagg tacttcatg atttacacat 120
45 ctgtgatatg cgtgtgtggc acgtgtgtgt cgtggatgt gtatctgtgg cgtgcataatt 180
tgtgggtgtgt gtgtgtgtgg cactgtgtgt tccatgtgtgt gtgtgtgtgt ggtgtgcata 240
tgtgtgtgtc tgtgacacatgt gcatgttcat gctgtgtgtc gcatgtgtgt gatgtgcata 300
tttgtgggtgt gtgtgtgtat gtgtccgtga catatgtgtc tctatggcat ggggtgtgtgt 360
ggcccttgg ctttactctt tccatccca ggcattggccc gcaacttgcgtt cttcaacgttc 420
50 tcgggtgtgtc gttttggggat cttccacatgtt agggccatccat ggtgtccctt 480
gtctgttcaac agggctgggc ctggagact gtaagccagg tttgagagga gagtagggat 540
gctgggtggta ctttctggaa cccctggcac ccccaaggacc ccagtcgtgc ctatggccgc 600
tccatggat ataggaaggc tgattcaggc ctgcgtcccc gggacacact cttcccaagag 660
cggccggggg ctttggggat cggcagggggt gaaaggggcc ctgggtttgg gttccacccc 720
55 agtggatcatg agcacgttgg agggatgttgc cttcaatgtc gtgcaggcc ggggtgcaga 780
ggtgaagaagat ttttttttttgcgttgcgtt gggagaggc acatgtggaa acccacaagg 840
accttttttctt ctgactttttt gatgtt 865
60
<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens
65
<400> 10
tgtgggatgt gttttcatgt gtgggatagg tggggatctg tgggattgggt ttttatgtgt 60
ggggtaacac agatgtcaag gcgagcttcc ttctgtgtgt gggctgtcag gtgtccaaac 120

```

agtttatttgg aggagacat atttcccttta gaactatggc cgggtttata gtaagttagg 180
 ggtgtggagg ctccccctgg gtccttcgtt ctgttttctc cactctgggg tcgtgtgggg 240
 cctgtgtgg tgggtggccg gtgggcaggc ctccaggcc tcccttggtt cattggcctg 300
 5 gatgtggccc tggctacgtt ccgtccctgg aattccctg cgagtggag gettcttc 360
 tttttttttt tttttttttt tgataacaga gtctcgctt ttttgcctt 420
 ggctggagggtg gtttggcggtg atcttggcgt actgcaaccgt gtgttcctg agtccaagca 480
 attctcttgc ctccggctcc caagtagctg gatattatgg cggccaccac catgtgact 540
 aatttttgta attttagtag agacgggtt tcccatgtt ggcaggctg gtctcgaa 600
 10 cctgaccta ggtgtatcctc ccacccggc ctcccaaagt gctggatga cagggtgtaa 660
 ccggccggcc cggccggagac tcgtttccctg cagttccctg gagatctgca gctatcgct 720
 cctggcggct tgggtgtggc aacctccctt ttcccttcctc aggttcctg agggtttttt 780
 ccattttatgtt actcttcata cagaagaggtt tcaatgttgc tgattttccg gctttttctt 840
 gcttaattgg tggatgtgtt ttatcgatgg cttcccttca tttcccttag gctttgttta 900
 ttgttggttt tccggcttca tgaaggaaaa gtttcgatgg tggatgtttt aactttttttt 960
 15 tctaaacaag catctgaagt tgccgttttc cctctaaagc aggatcccg aggcccctgg 1020
 ctgtggagggtg gcacccgggtt gggggcttggt aggaaccccg cgcacagccg gaggctagg 1080
 ggggggtgtggc gagccggcgtt tccggcttca gccccccccc tctcaatgttca gcaatggcat 1140
 ggggtgttca gaggccggcaca cacccactgtt agaaactgttgc gtggaggggg tctatgtttt 1200
 20 gtgtcttca tggaaatcta atgcctgttgc atctgggggtt gaaacggttt ctccaaacacc 1260
 catcccccttc cccactgtctg tccatgtggaa aaatcgctt ccacgaaaacc agtccctgg 1320
 accacaatgg ttggggaccc tggatctaaag acctgtttca gcaatggcttc gtcgtgtttt 1380
 atatattttggc ttttttttttgc tggatccatgg aatattatggg attttgcgttgc tgctttccgc 1440
 cgacccatgttca cccatgggtt attttggggc gtgttgcgttgc tccctgggtt gggagggtt 1500
 25 caggccccat gtatcttccat gttactgttcc tccagggttgc tttccagggtt gtaatcgatc 1560
 tcgatgtgtt tttagcccac ggcctggcg ccagcttctg gggggctgggg aacatgttca 1620
 agcacagagt cacccgtggc gtcttttgcgtt gcttcacaaag ctcggggctt cctgtgttcc 1680
 tggatgtgttgc tggatctgttgc tggatcttgcgtt gggggccggg gctttagcagg 1740
 tccctgtatgttca aatgacaagaatgg tggatggggggtggatgttgcgttgc tggatgggttgg 1800
 30 tctctctccgc gctgtttccat gacttttcttc tggatgtgttgc tggatgttgcac ctgcattttttt 1860
 gcaatccctc cagcaactggg ctggagaggc cccggaggtt gactgttgcact tggatgttgc 1920
 gactgtggat ggcgttgcgtt cacgggggttca tggatgtgtgg tggatgttgc tggatgttgc 1980
 tcaacagggtt ctgatgtgttgc tggatgtgtgg atgggggttgc tggatgttgc tggatgttgc 2040
 gtggatggcg gtcgtggggc tggatgtgttgc tggatgtgtgg atgggggttgc tggatgttgc 2100
 tggatgtgttgc tggatgtgttgc tggatgtgtgg tggatgtgttgc tggatgttgc 2160
 35 gggatgtatgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2220
 ggcgggtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2280
 gactgtggat ggcgggtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2340
 ctgatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2400
 cggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2460
 40 ctgtgtatgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2520
 ggtctgtatgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2580
 gtggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2640
 tctgtgtatgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2700
 45 ggtatggcggttctt ggtatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2760
 tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2820
 gggatgtatgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2880
 ggcgggtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 2940
 gtggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3000
 gggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3060
 50 gatggcggttctt gtttttttttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3120
 ggtatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3180
 acagggttctt gatgtgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3240
 gtggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3300
 tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3360
 55 gtcacagggttctt gatgtgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3420
 actgtgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3480
 tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3540
 tcggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3600
 acttttgcgttcc tccggccccc cggcccccgtt tcccaaaaca gaatgttccc aggccgttcc 3660
 tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc tggatgtgttgc 3720
 60 agtgcggccatc tctggccggg gcaatggccaca tttgtggcttc atgccttctc ctctggccgc 3780
 ag 3782

```

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens
5
<400> 11
gtctggcac tgcctgcag gttggcac ggactccag cagtggtcc tcccggc 60
aatcaactggc ctcatgaccc gagagactgt tgccctggg gggcagtggg gggaatgagc 120
tgtatgggg gcatgatgag ctgtgtgcct tgccaaatc tgagctggc catggcaggc 180
tgccacagct gtcgcattca ggcacctgt cacgtttgac tgcggggct ctctccagtt 240
ccgcgtggcc ttgttcatg atttgtctaa tgcgtttctt ccgcgttttgc atcttgggc 300
caaaggaaag gtgtccccctt cctttaggg ggcaggccat gtttggccg tgcctggcc 360
agctggggcc tcagtgcgg gtcggaggcc aaaggaaacg tgcctccctt cttagggaga 420
cgggccgtgt ttgagccacg ccccgctgg cgggccttc agtgcgtggg ctgtccacgt 480
ggccctgtgg cccttgcag atgtggtctg tccacgtggc cctgtggctc ttgcagatg 540
ccctgttagca ctgtgcggc tcttagggac agtcgtgtcc accgcattgg gtcagagac 600
ctctggggca atttcttgg cttccagggt ggggggggg gtttggggc ctgtggggc 660
ccagaccctg tgccggcagc actccctggg acatatggc atccggggcca 720
cggtgggtgt tggtgggtgt agccacgtg gacccacagg tggcccgag gagacgttct 780
20
gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcgcccg gccagccac 840
gtggggccctg catccacggc cagcccccga ctcatcaca aacactgacc cccaaaggga 900
cgaggggctt tgccacacgtg tgcctgcctg tttcagcacc caccggctca ctccatgtg 960
tctccgtct getttcgcag 980
25
<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens
30
<400> 12
gtgagtcaagg tggccagggtg ccattgcctt ggggtggctt gggcgccgt gcagggcttc 60
tgctcaccc tctccgcctt cttcccccact gncccttcgc cccggccac cagacttcc 120
tttctggcc cccggccccc cccgcttcgg ggttgcggc tcccgaggcc cggggaaat 180
ggctcggtt gggcggccgg ggggggggg gggccacac gggctggaa aatggcaagc 240
gggggtgtgg gtttgccttc cttggggggac gggggccggg ggggtgttctt gggcgggtt 300
tgccggcggc gtttgcgttgc gcacgttgc agtccaaatg tactactgac gtcggacacc 360
cggtcttcac acgtttgtat ctcttcctcc cgataaaaaa ggattttatac cgattctcat 420
tccctgtccctt gtcgtgtgac cccgggggg ggcggggctt ttctctctgt gactagattt 480
cccatcttggaa aagtgcgggg ttgaccgtgt agtttgcggg ttcggggggg cctgtgggtt 540
40
ccatggggca gggcccttc ggggggggg ggggggggg gggccacac gggccactca 600
cggtggtaga gccacagtgc ctggtgcac atcacgttctt ctggatttta agtaaaaacca 660
cacacccccc ggcaggccatc tgcctgcgac cctgtgtgtg cttggggaga gtttggtagcac 720
ggaggaaatctt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 780
ggggcccttc ctctggggatc gtcctccacgt gataaaaggac ttcggacacg ttcggaaatctt 840
tttattttttt aataataacta tttttttttt cttttttttt aataactaat ggtatcagca 900
attataatattt ttatattttttt attataatattt gttttttttt gttttttttt 960
cacaatttgc acatggcage agagttaattt ttggccgggg gacacgtgtg cacaatgtgtg 1020
taaggccccc ccaggccccac agaatttcgtt gacaaatgtt cttttttttt gttttttttt 1080
cggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
50
gtggcagggtt tttttttttt gttttttttt gttttttttt gttttttttt tttttttttt 1200
gggttactgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1260
tggttccatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1320
acagagagag tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1380
gggttccatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1440
55
gccccatctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
actcaactgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620
gctgagatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1680
gggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740
ctggcccttc cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1800
ctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860
gctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920
tcggcgtggca gccatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt 1980
actgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2040
60
gtcatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100

```


gcacctgtgc tctggggcatg gctgtgtcc tggaaacgttc cctgtccctgg ctggtcagggg 540
 ggtggccctcg ccaagaatcg acaactttat cacagagggga agggccaatc tggggaggcc 600
 acagggccag cttctgcctg gagtgcggc aggtgggtgc acaaggctcg ggctgttacc 660
 aaaggccgtg cgggcaccc agggccccggc ctcacccatca acaaggctcc cgagccactg 720
 ggagctaat gcccaggagc cgaaggccctc gccccatag ggtcgagaag gaggtgtgagc 780
 atttgtgtta ccggggcccg agggtgcgcg aattaccctg caacactgtat gtggaaatgg 840
 gtcgtcgctt atcgtggaaa cccagcaagg gtcacccggg gaggtttcca ttacaagggtc 900
 gtaccatgaa aatggttttt aaccggatgt cttgcgcctt catgtctgg cagggaggggc 960
 agagccacag ctgcatgtta ccgccttgc accagctcca gaggcttggg accaggctgt 1020
 ctcaagtccca ggggtgcgtcc ggctcagacc gcccctctt ctgccttctc tctgtcgctc 1080
 aaatctttccc tcgttgcat cttccctgacg cgtgcctggg ccctctgtca agctgttgc 1140
 ctcccttcgg gaaaaacccttg ggggtgtgc gatacagggtt ccactggagg ctggagggtt 1200
 ctgacactgt gggtgacccc agggtccagc tggcgtgtt ggggctctt tggccatgtc 1260
 tgaggtcaga ggagttttcc caggtaaaaa ctctgggaa actcccaaggg ccatgtgacc 1320
 tgccacccgtc tccctccatca ttcaagtcag tcttgccttc atttcccccac cagggtctct 1380
 agctcccgagg agctcccgagg gggggccctgg gtcacggca gggcggtgtga gtttcccccac 1440
 ccatgtgggg accctttgggtt agtgcgttgc tgggttagcc ctgaggaggc cgagatgcga 1500
 tggggccacgg gccgttccca aacacagatc caggcacgtg gaaggccccag gaatccccctt 1560
 ccctcgaggc aggagtgggg gaacggagag ctggggcccg atttacggc agccaggctg 1620
 cagtggggca gggtgtggg gtccacgtgg cgctggggc ggggtctgtat tcaaattccgc 1680
 tggggctccgg ctttcctggc ccgtgtgttgc cgccgcctcca caccggctgg ggggtggacgc 1740
 cccgaccctt agcagggtggc tattttctcc tttggaaagag agcccttcac ccatgtcttgc 1800
 tggttttctc ctgggttgcgg agcgtggggc tgggtcaacc cccggacattt aggttatttt 1860
 atttgtttaa aaacattctg ggcttgcgtt ccgttgcgttca aataatggggaa aaagacatcc 1920
 cacctcagca gagttactga gaggctaaaa cccgggtgtt ggcttgactg gtgtgtatctc 1980
 aggtcattcc agaagtggct caggaagtca gtgagaccag gtacatgggg ggctcaggca 2040
 gtgggttgaga tgaggatcac ggggggtttca ggcaatggggtggg acatgggggg 2100
 ctcaaggcact ggggtgagat aggtacacccgg ggggtcttggc cagagggtca gaccaggatc 2160
 acggggccctc tgatcacacg cacaatatag cacaatgtca catgtgttgc ttcatgttgc 2220
 ccaggctctgt gcacacccgtc cccaaaggcc caggaagctg agaggccaaa gatggagggt 2280
 gacaggccgtc ggcgggtggc tcacacctgt agtcccagca cttrgggagg ccgaggcgag 2340
 aggatccctt gagcccaaggaa gtttaagacc agccgtggca acatagtaga accccatctc 2400
 tatggaaaat aaaacaaaaattt attagctgaat catgtgggttgc tggcgtctgtat gttccaaatac 2460
 ttggggggat gaaatggggat gatcaatgtca gcccaggagg tggaaatgtc agtggactgtc 2520
 gattgcacca ctgtactgtca gcctgggtga cagagtggaa gcccataatca aacaacacaa 2580
 agaagactga caaatgcagt ttcttggaaa gaaaatcta gttagaaactt aacctacacaa 2640
 cagaagccaa gtcgggttgc cgggttgcgtt gagatggat gatgggttccat cacaccatca 2700
 ccccaagaccc aggggtttatc caccacaggg gcccgggtgtt cagaagggttgc ggcggacgc 2760
 ttgtatatacg atgacatcaa ggttgcgttca cgaaggccggc gattatgtat gatgttgc 2820
 tgggtatcacaa ggaacaatggg ataaaacttggaa aaccccttggg gcccctccggg aacaggggct 2880
 aattcagaagc cagcatgggg ggctggcgttcc caggatgggg ctgtttccatgc ctccacatgc 2940
 gtgttccatc agatgggttgc cagaaacgtca gtgttccatgttgc acacacacag acacgcagct 3000
 acttcgcacac acaaggacac acacacacat gcatgtatgc atccgtgttgc gtgcacccgt 3060
 gcccattgggg aaacccatgtc atgtgttgcgttcc atgcacggcac acaggccacccg gtggggccat 3120
 gcccacaccc acgagccaccc tctgttgcgttcc tctgtatgtc tccgcacccatcc 3180
 tcttcag 3186

<210> 18
 <211> 781
 <212> DNA
 <213> Homo sapiens

<400> 18

55 gtatgtgcag gtgcctggcc tcagtggcag cagtgcgtgc ctgtgtgtt tagtgtgtca 60
 ggagactgtg tgaatctgg cttagggaaat tcttacccctt tttcgatcatc ggaatgggtt 120
 taacccaaacc actgttcaggc tctgtgtccccc gcccctctgtt ggggtggatca gaggacttgc 180
 tggaaaggagc agggactgttc tggggatgttc catctttccccc accttgcgttcc gcttggggaa 240
 ggcgtggggg gcctgggttcc tctgtttgtc cccatgttgg gatttggggg gcctggccctc 300
 tctgtttgtc cctgtgttgg gatttggggtgc tctccgttcc atggcaatcta gggcccttgc 360
 gcaaaaaaccgg gccaagggtt tagggaggagg ccaggccccc gtcacccccc cccttccttgg 420
 agcagaggcc gctgtatccat acggacagacg cccggccgtt cctctgttcc ccagtccaccc 480
 tctctgtcccc ctggacaccc tggccatgttgcgttcc tctgtatgttgc tggaaatcc 540
 agccatgtcg aacatgtcggtt cctgtatgttgcgttcc tttttgttcc 600

gtggaaatcc cacctggaga agccgaagaa aacattctg tcgtgactcc tgcgggtcctt 660
 gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagcttccg 720
 gtgtctctcg ggaggggagc tgggctggc ctgtgactcc tcagctctg tttttccccca 780
 g 781

5 <210> 19
 <211> 536
 <212> DNA
 <213> Homo sapiens

10 <400> 19
 gcaagtgtgg gtggaggcga gtgcggggccc cacctgccc ggggtcatcc ttgaacgccc 60
 tgtgtggggc gaggcgcctc agatgtgtc gaagtgcaga cgccccggg cctgaccctg 120
 ggggcttggc gccacgctgg cagccctatg tgattaaacg ctgggttccc caggccacgg 180
 15 agcctggcag ggtccccaaac ttcttgaacc cctgtttccc atctcagggg cgatggctcc 240
 ccacgcttgg gagccttctg accccctgacc tigtgtccctc cacagcttcc tccctggctg 300
 ctggccctgag ctccctggggt cctgagaacg ttcttcccccc gccccccgc tccacgtca 360
 ctgggtcgc tigtctgtcg ccccccgtggg ggggtgtctg tcccttaact gaggttccca 420
 20 ccagccaggc ccacgaggc caggccctgc ctggccggcc accccacacgt ccttaggaggg 480
 ttggaggatg ccacctctgg cctcttctgg aacggagtc gatttggcc ccgcag 536

<210> 20
 <211> 3179
 <212> DNA
 <213> Homo sapiens

25 <400> 20
 atctcatgtt tgaatctcaa ttttgtcaatgc atagacacca ctgtatgcaat ttacagaage 60
 ctgtgagtga acgggggtgtt ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
 30 gtctatgatgtt gaatgggggtt gtggtcgttg cggggccatg gcctgtgtgg gcctggagg 180
 tttctgtatgc ttttgtggcag gaggggaaagg agggtagggg atagacatgtt ggagccccca 240
 cccttggaaaga cataaacatgtt agtccaggcc cgaaggccag cagggatgtt gggggcccaag 300
 ctggggccgc ggggatgtt gggggatgtt ccagggtggc agggatgtt gggggcccaag 360
 ctgggggtgtt gggggatgtt gggggatgtt gtctgggtgg cggggaaatgtt gggggcccaag 420
 35 ggctggccccc ctccctccccc tgcctccccc ctgcaggccgtt gatccggat gtgttccctt 480
 ggtgcacatc ctctggccca tcaatgttca tggaggtggg gggcaggggc atgacacccat 540
 cctgtataaa atccaggatt ctcccttctgg aacggcccaaa ctcaatgttca aagtccacatt 600
 ccgccttctgg ccatttctttt aagagtagac caggatgtt atcttgcgtt ggtgggtttagg 660
 gtggggcgtt ggaggggtgtt gacacaggcc gcttcagggtt ggggtgtttagg atgtctcttc 720
 40 atcccttat catctccccc ttcatcttctt catcttcttca tcatctccca gtctcatctg 780
 tttctctttt atctcccaat ttcatctgtt atcccttttac catctcccaat ttcatcttctt 840
 tatecttta ttcctcttagtca tcatccagac ttacctccca ggggggtgtc caggctcgca 900
 gtggagctgg acatacgcc ttcctcaggc agaaggaaat ggaaggattt cagagaacacg 960
 gaggggccgc tcagaggccatc gcaatgtt ggtgttggaaat cagcccccctt tcagaagttt 1020
 45 gttttggccca cacaacccatc aggcccttgc gtgagttgtt ccagggccctt ccagcaggcc 1080
 ctgtggggggg ctcttatgttca tggccgggttcc tctatgttgc cacatggcagc agggcttctg 1140
 gttttagtgc agcccgccatc tggctgttgc tgggggtgggg gctttagtgc actggatatg 1200
 gcttcattttt ttgtctgttc ttcagagaaat gtctgttgc ccgcgttccat tttttttttt 1260
 gggcccaatgtt ccacagactt tttttttttt tttttttttt gttttttttt gttttttttt 1320
 50 agtctgttgc aaggaggggcc ttttggccatc cggccctgggg ggccttgc cctgttccat 1380
 ggaaggggccatc gggcccccggc gccgtggggc gacgacccatc agtggaggtt tggcacaac 1440
 agggccggggatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500
 gaccaacaggccatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1620
 55 tagacccatcaaaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt 1680
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740
 gttttagtgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1800
 gatccctccggccatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860
 ctggcactttaaaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt 1920
 60 ctagtttttgc ggaaggccatc tttttggccatc cggccctgggg ggccttgc cctgttccat 1980
 gggtaacatcaaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt 2040
 ccagcatctgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100
 ggtttagtgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
 gaccctgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220
 gaaggaaagaatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280

5 taggttagact gtcaaatctc agagcaaaat gaaaataaca aagtttaaa gggaaagaaa 2340
aaccggcgt cttggactt ccttaggcgt gaacttcatt tcaagcgact tcctccaca 2400
gacaaggcgta tatggagcga gtgagtcaa agcagaagg gaggagaagc aggcaagggt 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaagggg gtgggtttt tcctgcctca 2520
gccccacgct cctgcccgtc ctgcacccgtc tgtaaccgtc gatgtggtg ccaggtgccc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttt gtgggtttca gaagccccag 2640
gcacttgtgg caggcacaat tacagccct ccccaagat gcccacgtcc ttctcctgg 2700
acctgtgaat gtgtcacccg caaggcagag gttggtgaag gtcgcagggtg gaatcacggc 2760
tgccagtcag cegatcttaa ggtcatacttg gattatctgg tggggctcgat atggccacaa 2820
gggtccctag aagtggagaga gggaggcagg ggagagttag agaggggacg tgagaaggac 2880
caactggccac tgctggcttt gagatggagg agggggctcc cagccaaggaa atggggcag 2940
ccgctccatg ctggaaaagc aagcaatctt ccccggtctt gagggcacac ggcctgccc 3000
acgcctcgat ttcaaggccag tgggacctgt ttcaagcttc cggcctccag agctgtaaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
10 15 cagtagcaggg aaatgaatac agggacagt ctcagatgtca ctctcagccc acccctggg 3179